STATISTICS

Textbook

BASIC STATISTICS

For

INTERMEDIATE CLASSES

PART-I

by

Muhammad Saleem Akhtar
Ghulam Hussain Kiani

MAJEED BOOK DEPOT
Lahore-Faisalabad-Rawalpindi

BASIC STATISTICS

(A Text Book for Intermediate Classes)

PART - I

By

MUHAMMAD SALEEM AKHTAR
Department of Statistics
Govt. Gordon College,
Rawalpindi.
0333 - 5127797

GHULAM HUSSAIN KIANI
Ex. Associate Professor of Statistics

MAJEED BOOK DEPOT
22-Urdu Bazar, Lahore
Ph:042-37311484, 37355187

Aminpur Bazar, Faisalabad. 041-2412641
6th Road, Rawalpindi. 051-4423948

REVISED EDITION 2012-2013

Author Muhammad Saleem Akhtar
0333 - 5127797

Publishers: Kashif Mukhtar
Majeed Book Depot
Urdu Bazar, Lahore.

Printers: Z.A. Printers, Lahore.

Copy: 500

Composing: Shahid Ayub Channa

Price: Rs. 225/-

Preface

Basic Statistics has been written to serve as the text in the Intermediate course in Statistics for the students of Part-I. The book has been written according to the new syllabus approved by the National Bureau of Curriculum and Text Books, Ministry of Education, Govt. of Pakistan, Islamabad. Thus the book would be equally useful for students of all the Education Boards in Pakistan. It has been written especially for Intermediate students but the students of B.Com; M.A. Economics, M.Sc. Psychology, M.Sc. Geography, Business Administration or any other area of social sciences can also benefit from this book. For beginners interested in learning the basics of the subject, this book should indeed prove to be a valuable gift. We would not like to make tall claims at this stage. We would rather leave it to the subject teachers and the students to evaluate the book and form their own independent opinions. We, on our part, are glad to say that we have fulfilled a long outstanding demand of our ex-students for such an undertaking.

The book is the product of combined efforts of the youthful energies and long teaching experience of B.Sc, B.S and M.Sc. classes of one author and long teaching experience of the other. Both the authors are proud of their sustained and successful efforts in bringing out a book dedicated primarily to the problems faced by the subject teachers as well as the students. Some typical features of this book are:-

(i) We have stressed on the solved examples so that the students may easily learn the subject by studying the examples.

(ii) Theory has been written in simplified language. It would seem as if the teachers were speaking to the students and the simple situations of practical life were being explained.

(iii) Diagrams have been used for effective illustration.

(iv) Exercises contain some questions which are different from the age-old pattern of questions. New questions have been added to invoke fresh thinking. Students have been encouraged to generate their own exercises for better understanding and assimilation of the subject.

(v) Short Definitions, Multiple Choice Questions and Short Questions are also included which are according to new syllabus and which will be helpful for students to have a command on the subject.

Lot of pains have been taken in proof reading but we do not claim that the book is free of typing/printing errors. There may be errors on the part of the authors. We apologize for any inconvenience caused on this account. We shall be grateful if any such errors or omissions are brought to our notice. For the improvement of the book, we would welcome any suggestions or criticism particularly from our friends in the teaching community.

We are thankful to our relatives, friends and colleagues who encouraged us for writing this book. We have special words of thanks and appreciation for Muhammad Khurshid Khan, the computer operator who worked with us with devotion.

Majeed Book Depot, Urdu Bazar Lahore, our publishers, have the credit of bringing out the book in time despite acute limitations. We owe extra special thanks for them.

September 1, 2012

Muhammad Saleem Akhtar
Ghulam Hussain Kiani

CONTENTS

Chapter 9

Chapter 1

INTRODUCTION

1.1. HISTORY

Statistics is a very old word. It is from the English language. Nothing is known about the exact origin of the word. It is written in the books on statistics that this word perhaps been derived from some other language. For example there is a word *Status* in the Latin and in the Italian there is a word *statista*. Another word is *statistik* in the German language. All these words mean "the political state". Shakespeare used a word statist in his drama Hamlet (1602). In the ancient times the statistics was used by the rulers. The application of statistics was very limited. It was nothing but the record of armed forces, labour force, agricultural land, area and population etc. The records of human beings were very important to the kings. There is a lot of gap between the old statistics and the modern statistics. But the old statistics also makes a part of the present statistics. The English writers have used this word in eighteenth century in their works. W. Hooper used the word *statistics* in 1770 in his translation of Elements of Universal Erudition, a famous old book written by Baron B.F. Biefeld. In this book one Chapter is on statistics, which has been defined as "the science that teaches us what is political arrangement of all the modern states of the known world". E.A.W. Zimmermann has defined the word *statistics* in 1787 as "that branch of political knowledge which has for its object the actual and relative power of the several modern states, the power arising from their natural advantages". Statistics has developed gradually during the last few centuries. Most of the work in statistics has been done in the end of the nineteenth and the start of the twentieth century. In 1875, a German physicist F.R. Helmert (1834 – 1917) developed a method called *chi-square*. This method brought a big revolution in statistics.

In 1900, Karl Pearson (1857 – 1936) worked on the *chi-square* and suggested a very useful test called *goodness-of-fit test*. Sir Francis Galton (1822 – 1911) worked on Regression and suggested the techniques which are used in modern statistics. A very famous and useful method of Analysis of Variance (ANOVA) was developed in 1923 by Sir R.A. Fisher (1890 – 1962). The modern statistics in its present form is not only helpful to the rulers, but is also helpful to the experts working in different fields of life. Knowledge of statistics has become important for a common man. A person without a reasonable knowledge of statistics is not a literate person. Modern statistics is being applied in various fields of life. It increases our wisdom to understand the matters of practical life. A man equipped with the knowledge of statistics is a better ruler, a better policy maker and a better administrator.

1.1.1. OBSERVATION

In statistics, numerical measure of some condition is called an observation. "Mr A is a rich person". It is not an observation because the income of Mr. A has not been mentioned numerically. Income of Mr. A is Rs.6000 per month is an observation. Mr. B is intelligent, he is hardworking, he is very kind, he is a good player, he is honest. None of these statements is an observation. The work on statistics may be called statistical work. The statistical work starts with a set of observations. The observations may be about the wages of workers in factories, about prices of commodities in a market, about marks of students, about the land holdings of different farmers, the number of accidents per day on the roads, the daily production of a factory and so on. Thus we start with observations. Observations are the take off point for statistical work. Observations serve as raw material for statistical work. We take the observations by counting or by measurement. The number of patients in different hospitals are counted and their blood-pressures are measured.

Weights of commodities consumed, weights of students, neck sizes of students, chest sizes of players are all result of measurements. Measurements are never exact. A weight of 60 kg. means that the weight lies between 59.5 and 60.5 kg. A weight of 60.5 kg. means that it lies between 60.45 and 60.55 kg. The number of students in different sections are counted. Counting generates exact figures.

1.1.2. POPULATION

The word population or statistical population is used for all the individuals which possess some common characteristic and we are interested to know some properties of those individuals. The individuals may be called the *study units*. The study units may be living or non-living things. We may have to study a population of fans produced in a factory during certain period, we may have a population of fruit trees in a garden, a population of children in primary schools, population of patients in hospitals, population of beggars in a country, population of salaried persons in a country, population of fish in an ocean, population of cows in a country. A population may be finite or infinite. It is called finite if its individuals can be counted. For example the population of income-tax payers, the population of smokers in a country. The population is called infinite when its individuals cannot be counted. For example the number of sun-rays and the number of dust particles on the surface of earth. The number of fish in an ocean may be very large but countable. The number of stars in a galaxy may be countable. Such populations are called countably infinite (very large but countable). The size of the population is denoted by N. If we have 500 individuals in a population, we write $N = 500$. The size of the population is usually very large.

1.1.3. SAMPLE

Sample is any part of the population which is selected from the population. The sample is studied and on the basis of the sample study, we try to know something about the population. This journey from the sample to the population is called inference which is an important branch of statistics. The method of selecting sample from the population is called sampling. The sample size is denoted by n. The ratio

$\frac{n}{N}$ is called sampling fraction. Sample is studied to know the properties of the population. The sample study saves time and finances and in general the sample study provides results of better quality. In statistical studies, most of the work is done on sampling basis. The medicines, seeds, fertilizers and other products are first prepared on small scale and after approval, they are manufactured on large scale. In subsequent Chapters, we shall make many numerical calculations. All calculations will be about samples. Whenever there is a calculation about the population, it is mentioned clearly.

1.1.4. PARAMETER AND STATISTIC

Any measure of the population is called a parameter. Suppose we are interested to know the percentage of smokers in our country. This percentage pertains to the population of smokers and is called parameter. It is usually unknown. The symbol μ(mue) is used for population mean and the symbol σ^2(sigma-square) is used for population variance. These symbols pertain to the population and are called the parameters. The parameters are usually unknown and are estimated through samples. Population parameter is a fixed quantity and may be called a constant.

Any measure of the sample is called statistic. The statistic depends upon the sample. Different samples have, in general, different values of the statistic. The symbol $\bar{X}$ is used for sample mean and S^2 is used for the sample variance. These are the statistics (plural for statistic). The sample statistics are used to draw conclusions about the population parameters. The process of drawing these conclusions is called inference or inferential statistics.

1.2. MEANING OF STATISTICS

The word statistics has three different meanings which are discussed below:

(a) Statistics in plural sense

The word statistics in plural sense are the numerical observations collected for some definite purpose regarding some field of study. These observations may be for the sample or the population. These observations are also called data. The observations about the wages of workers are called the statistics of wages. The prices of commodities taken from the market are called statistics of prices. It is usually said that the statistics of education and agriculture are poor in our country. This statement means that we do not have the correct information about the school-going children, about the failures, about the drop-outs, about the results and about the unemployed educated persons. We have poor agricultural statistics because we do not have correct figures about the production and consumption of different crops, fruits and vegetables. One can talk about one's personal statistics e.g., age, weight education, income and children etc. We may say "he is a man of good statistics". It means that he has a good fortune. He has wealth, house and other facilities.

(b) Statistics in singular sense

Statistics in singular sense is a body of methods used in the collection, presentation, analysis and interpretation of data. The word *scientific method* is also

used for the steps which are taken to study the data. This meaning of the word is close to the word statistics as a subject.

(c) Statistics as plural of Statistic

Statistic is any measure of the sample. The proportion of smokers in a sample may be denoted by $\hat{p}$. It is a statistic. We can find the mean $\bar{X}$ of the sample. It is also a statistic. When we have more than one statistic, we shall use the word statistics as plural for statistic. The sample mean $\bar{X}$ and the sample proportion $\hat{p}$ are the sample statistics.

1.3. CHARACTERISTICS OF STATISTICS

The word Statistics in plural sense has some characteristics which are discussed below.

(i) Statistics are aggregate of facts

A single observation does not make statistics. The income of Mr. A is Rs.20000/- per month is not statistics. The observations must be in a group expressed in proper units if applicable. The marks of some students may be 20, 30, 40, 50. There is no unit of measurement. But if we talk about prices of meat in different markets, we must write the units i.e., 'Rs.' or '£' with the observations. The prices like 50, 55, 60, 62 do not make statistics. The figures as Rs.50, 55, 60 and 62 are statistics provided they have been collected for the purpose of some study.

(ii) Statistics are numerically expressed

Only numerical statements make statistical data. The statements like "he is a good player" or "he is a poor person" are not the numerical statements. Income of Mr. A is Rs.8000 per month and that of Mr. B is Rs.9000 per month are the numerical statements. These numerical statements are the statistics of the income of Mr. A and B.

(iii) Statistics are affected to a marked extent by multiplicity of causes

Every observation is the net result of many forces. There are many forces which exert their influence in the field from which the observation is to be taken. The marks of a student depend upon his hard work, his intelligence, the intelligence of his teachers, the devotion of his teachers, the financial and the educational level of his parents, the locality in which he is living, the transport facilities and many other factors. Some of these forces are more powerful than others. The marks of a student may be high due to his hard work or due to his intelligence. The marks of some other student will be affected by another set of multiplicity of causes. Thus the observations (statistics) are affected to a great extent by many forces.

(iv) Statistics are enumerated or estimated according to a reasonable standard of accuracy

The measurements are never exact. There is nothing like exactness in nature. Suppose we take an observation on the weight of a player. His recorded weight is 70 kg. Is it exactly 70 kg? No it is never exactly 70 kg. The weight lies somewhere

between 69.5 to 70.5 kg. It is only for practical purposes that we have recorded the weight as a whole number. We may make the observation on weight and our observation may be 70.2 kg. Is it exactly 70.2 kg? No, it is somewhere between 70.15 to 70.25 kg. When we are interested in the weights of the players, it will be meaningless to take the observations up to 4 places of decimals in terms of kgs. Sugar consumed by consumers on daily basis may be measured in gms. The amount of iron in the body of human beings may be measured in milligrams. Consumption of meat by consumers per day will not be measured in tons or milligrams. The suitable measure will be 'kg'. The roads are not measured in cms and the length of rooms is not measured in kms. We take the observations according to a reasonable level of accuracy. When we make a measurement, we get an estimated figure and not exact figure.

Enumeration means counting. The population of big countries is counted in millions, the population of small cities may be counted in thousands and for the population of villages the actual figures may be used. It may be noted that in the counting process the exact figures are possible but in measurements the exact measurements are never possible. Depending upon the situation, we use suitable units of measurement or counting.

(v) Statistics are collected in a systematic manner

A systematic method of collection of statistics is adopted. Suppose we want to carry out a study about the quality of primary education in our country. First we shall make a list of all the primary schools, a sample of some schools will be selected. In the selected schools, we shall make lists of the students and teachers. The qualifications of teachers and the ages of students will be recorded. The scores of the students in the house examinations will be taken. All this work is to be done in a systematic manner.

(vi) Statistics are collected for a pre-determined purpose

The mere figures which do not serve any purpose are not statistics. Those figures make statistics which have been collected to carry out some study, survey or inquiry. When we take the price figures from the market to see as to how much increase/decrease has taken place in the level of prices, then the price figures are called statistics. These figures are to be used for a definite object in mind.

(vii) Statistics must be comparable to each other

Observations taken from one place should be comparable with the observations taken from some other place. Suppose we have to carry out a study about the literacy level in different provinces of our country. In the Punjab the definition of literacy may be different from the definition of literacy in the Baluchistan. If the definitions are different the statistics of literacy will not be comparable. Similarly in a study on housing facilities, a house in the N.W.F.P. may have a different meaning from a house in the Punjab. A farmer of the Sindh may not compare with a farmer of the Baluchistan. Statistics taken at different times should also be comparable. In the 1972 population census of Pakistan, the definition of a literate person was "a person who can read and write a simple paragraph in any language" and in 1981

population census the definition was "a person who can read or write a simple paragraph in any language". Obviously the definitions were different. Thus the statistics of literacy in 1972 and 1981 are not comparable. The definitions should be consistent.

Horace Secrist has given a comprehensive statement which includes all the characteristics described above. His statement is *"statistics are aggregate of facts, affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".*

The above statement is a detailed definition of statistics in plural sense but this definition does not include the singular meaning of the word statistics. Thus this statement is not the complete definition of the word statistics.

1.4. DEFINITIONS OF STATISTICS

Statistics like many other sciences is a developing discipline. It is not something static. It has gradually developed during the last few centuries. In different times, it has been defined in different manners. Some definitions of the past look very strange today but those definitions had their place in their own time. Defining a subject has always been a difficult task. A good definition of today may be discarded in future. It is difficult to define statistics in a clear and acceptable manner. Some efforts which have been made in the past for defining statistics. Some of the definitions are reproduced here:

(i) The kings and rulers in the ancient times were interested in their manpower. They conducted census of population to get information about their population. They used this information to calculate their strength and ability for wars. In those days statistics was defined as

"the science of kings, political arithmetic and science of statecraft".

Today this definition may seem very strange but it is true that our Govts. use statistics. All the future planning is based on statistics.

(ii) A.L. Bowley defined statistics as

"statistics is the science of counting".

This definition places the entire stress on counting only. A common man also thinks as if statistics is nothing but counting. Statistics today is not mere counting of people, counting of animals, counting of trees and counting of fighting forces. It has now grown to rich methods of data analysis and interpretation.

(iii) A.L. Bowley has also defined it as

"science of averages"

This definition is very simple but it covers only some area of statistics. Average is very important in statistics. Experts are interested in average death rates,

average birth rates, average increase in population, average increase in per capita income, average increase in standard of living and cost of living, average development rate, average inflation rate, average production of rice per acre, average literacy rate and many other averages of different fields of practical life. Large number of experts are involved in the calculation of different averages. Thus 'average' is an important calculation. But statistics is not limited to averages only. There are many other statistical tools like measures of variation, measures of correlation, measures of independence etc. Thus this definition is weak and incomplete and has been buried in the past.

(iv) Prof. Boddington has defined statistics as

"science of estimates and probabilities"

This definition covers a major part of statistics. It is close to the modern statistics. But it is not complete because it stresses only on probability. There are some areas of statistics in which probability is not used.

(v) A definition due to W.I. King is *"the science of statistics is the method of judging collective, natural or social phenomena from the results obtained from the analysis or enumeration or collection of estimates"*. This definition is close to the modern statistics. But it does not cover the entire scope of modern statistics. Secrist has given a detailed definition of statistics in *plural sense.* His definition is given on the previous pages. He has not given any importance to statistics in *singular sense.* Statistics both in the singular and the plural sense has been combined in the following definition which is accepted as the modern definition of statistics.

> *"statistics are the numerical statements of facts capable of analysis and interpretation and the science of statistics is the study of the principles and the methods applied in collecting, presenting, analysis and interpreting the numerical data in any field of inquiry".*

1.5. VARIABLE AND CONSTANT

The word variable is used for something which can take different numerical values for different individuals or objects. On different shops the prices of apples may be Rs.30, 32, 30, 35, 36, 32 per kg. This set of numerical values is called the *variable.* If apples are being sold at Rs.32 on all the shops, then we have a fixed price and we say that the price is *constant.* It does not change from shop to shop. Different patients may have different blood pressures but they may have same level of temperature. Thus blood pressure is variable and in this example temperature is constant. Different workers may have different wages but their medical allowance may be fixed. Thus wages are a variable and medical allowance is a constant. For variables we shall use the symbols x, y and z and for constants we shall use the symbols 'a', 'b' and 'c' etc.

1.5.1. DISCRETE VARIABLE

A variable is called discrete if it can take only some selected values in a given interval. Suppose in a certain town, the minimum number of family members is 1

and the maximum number of family members is 7. No family will have 1.6 members. The number of family members must be 1, 2, 3, 4, 5, 6 or 7 and no other value between 1 and 7. There is a jump from 1 to 2 because we can think of any number of values between 1 and 2. This variable cannot take any value between 1 and 2. This variable is discrete and can take values with jumps. When we select 5 bulbs from a factory, the number of good bulbs may be 0, 1, 2, 3, 4 and 5. This is a discrete variable. If 30 students appear in an examination, the number of passing students may be 0, 1, 2, ..., 30. It is a discrete variable. Discrete variable is generated by counting or enumeration. It may also be called count variable. A set of observations on discrete variable is called discrete data.

1.5.2. CONTINUOUS VARIABLE

A variable is called continuous if it can possibly assume all values in a given interval. Suppose the minimum temperature of a city in summer is 35°C and maximum temperature of the season is 45°C. The temperature can take any value between 35°C and 45°C. It may be 35, 35.01, 35.02, 35.1, ..., 45°C. The observed values of temperature on different days may be 35, 41, 42 etc. but as far as possibility is concerned, the variable can take any possible value in the given range. Examples of continuous variable are heights of students, amount of rainfall on different days, the consumption of sugar per family etc. A continuous variable is generated by measurements. It may also be called a measure variable. A set of observations on a continuous variable is called continuous data.

1.5.3. QUANTITATIVE AND QUALITATIVE DATA

A set of observations generated by counting (enumeration) or by measurements is called *quantitative data*. Both discrete and continuous data come under the quantitative data. The data on the number of accidents on roads, the number of teachers in different schools, the intelligence quotients of the students, the temperature in different cities, the consumption of electricity in different houses are quantitative data.

The word qualitative data is used for that information which is generated by observing the presence or absence of some quality (called attribute) in individuals. The individuals may be the students and the quality may be the intelligence. We examine whether or not intelligence is present in the students. The observations are recorded with the help of 'yes' for intelligence and 'no' for non–intelligence. Thus 'yes' or 'no' generate the observations and the data thus obtained are called qualitative data. Some examples of qualitative data are given below:

(i)	Number of intelligent students	Number of non-intelligent students	Total
	60	40	100
(ii)	Number of skilled workers	Number of unskilled workers	Total
	3000	2000	5000

1.5.4. ERRORS OF MEASUREMENT

It is possible to get exact observations in a process of counting whenever counting is possible. The number of students in different class rooms may be 30, 35, 46 and so on. These are exact observations. But if we measure the weights or heights of these students, the observations will serve our purpose but they will not be exact. They will be approximate. The reason is that there are the limitations of the observer (may be human being or some instrument) and the limitations of the measuring instruments. An observation may be in a whole number as 60" height, it may be up to one place of decimal as 60.2. It will not be absolutely exact even if it is measured to any number of decimal places. We can increase its accuracy by increasing the number of decimal points if possible on the measuring instrument but the exactness can never be achieved. In fact, we never need that absolute exactness which is only something imaginary and does not exist in nature. The difference between the exact value and the observed value of the observation is called error of measurement. It is always there whenever a measurement is made. It can be reduced but it cannot be eliminated completely. These errors are of two types namely biased and unbiased .

(i) Biased errors:

If the observations tend to be higher or lower than the exact observations, the errors are called biased. Some observer may intentionally take wrong observations in which case bias is called intentional or he may not be capable enough to make the measurements. He may be inclined in one direction. A teacher may award marks to the students less than what they deserve. It is called one-directional bias. The errors involved are also called systematic or cumulative errors. These errors are not reduced by increasing the number of observations.

(ii) Unbiased errors

Unbiased errors occur when the observer makes the observations some of which are higher than the exact observations and some are below the exact observations. Some errors are positive and some are negative with the result that they cancel out each other when they are added. These errors are, therefore, called compensating errors. They are also called random or accidental errors.

1.6. BRANCHES OF STATISTICS

The science of statistics can be divided into following branches:

(i) Theoretical statistics, (ii) Descriptive statistics,

(iii) Inferential statistics, (iv) Applied statistics.

1.6.1. THEORETICAL STATISTICS

This branch of statistics deals with the formulas and rules which are used in statistical work. These rules are used in *descriptive* as well as inferential and applied statistics.

1.6.2. DESCRIPTIVE STATISTICS

This branch of statistics covers the methods and principles of collecting observations, their condensation and finding the salient features of the data. It also

covers graphic presentation. Some basic calculations are made on the data to find out its properties. Suppose we make a graph of the imports of Pakistan during the last 10 years. We can read the graph to see as to what has been the growth of imports in the past. This study belongs to *descriptive statistics.*

1.6.3. INFERENTIAL STATISTICS

This is the branch of statistics which deals with procedures of drawing inferences about the population on the basis of the information obtained from the sample. This is a very important branch of statistics. The knowledge of probability is used to make inferences about the population with the help of the sample study. Suppose we want to have an idea about the percentage of smokers in our country. We take a sample from the population and the proportion of smokers in the sample is calculated. This sample proportion with the help of probability enables us to make some inference about the population proportion. This study belongs to the area of *inferential statistics.*

1.6.4. APPLIED STATISTICS

The statistical knowledge is used in different fields of life. It is used in the fields of business, economy, biometry, demography, agriculture, education, banking, insurance etc. The statistical wisdom helps us to understand the data in various fields of life. Suppose we take the prices of various commodities from the market. We can calculate the index numbers of prices to see the changes in the level of prices. This is a very important study for the consumers, for the businessmen and for the Govt. Such studies fall in the area of *applied statistics.*

1.7. FUNCTIONS OR USES OF STATISTICS

(i) It presents facts in a numerical form

An important feature of statistics is that it deals with numerical statements. "We had heavy rains this year". This is a statement which may be interpreted in different ways by different persons. Let us make another statement. In this year the rainfall was 90 cm. as compared to 70 cm. during the last year. The second statement gives a clear picture of the amount of rainfall. This statement contains figures which are statistics.

(ii) Statistics simplifies complex mass of data

The size of data is usually very large. It is difficult for the human mind to make out something out of the huge amount of data. The data are converted into certain calculations like totals, averages and percentages etc. Suppose we have data about human population divided into males and females, literate and illiterate, urban and rural, children and adults and so on. There is so much of information in this data that it is difficult to make some opinion about such a huge data. When the data are summarised by certain calculations, it becomes simple to understand it.

(iii) Comparison of data becomes easier

Sometimes different data are to be compared with each other. This comparison is not possible without the statistical tools. We apply some statistical methods which help us to compare these data.

(iv) Statistics studies relationship among different facts

Sometime when a variable changes, it produces a change in other variables. Statistics helps us to carry out studies to find the relationship between two or more than two variables. It has been observed that when prices of oil rise, the prices of fruits and vegetables also rise. This indicates that relationship between two variables exists. Statistics helps in studying such relationship.

(v) Statistics studies changes in a variable

Changes in a certain system are studied by statistical methods. The population changes from time to time. The prices of different commodities change with the passage of time. There exist statistical methods for measuring these changes.

(vi) Statistics helps forecasting

With the help of the available data, a fair idea can be formed about the future values of the variable. This is called forecasting.

1.8. LIMITATIONS OF STATISTICS

Statistics is applied in different fields of life. But it is not free from deficiencies. Some of the limitations of statistics are :

(i) Laws of statistics are applicable on group of observations. Any single observation even if it is very important to a common man cannot be dealt with by statistical methods.

(ii) Statistical methods are best applicable on quantitative data.

(iii) Some errors are possible in statistical decisions. Particularly the inferential statistics involves certain errors. We do not know whether an error has been committed or not.

1.9. IMPORTANCE OF STATISTICS IN DIFFERENT FIELDS

Statistics is applied in all those fields in which data are involved. Data are studied with the help of statistical methods. Data analysis requires an expert knowledge and a common man without the knowledge of statistics cannot properly analyse the data. Good data may be spoiled by a person who is not expert in statistics. Application of statistics is very wide but we discuss here some important fields in which statistics is commonly employed.

(i) Business and Economics

Statistical methods are widely used in the study of economic problems and the business activity. The relationship between supply and demand is studied by statistical methods and the decisions are supported by the observed data. The increasing prices are a big problem of business activity. The changes in prices cannot be studied by a common man. The imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics. The present level of prices is compared with the level of prices in the past. It helps a common man to have an idea about the level of prices in future. The wages of workers are linked with the level of prices. The problems of prices and wages are

very serious in our country. To study these problems, a thorough knowledge of statistics is required.

(ii) Banks

Statistics is very important to the banks. The banks do business with the help of deposits in the banks. The banks do not keep all the deposits with them. They do business with it and lend the money to the borrowers. The banks work on the assumption that all depositors do not withdraw their deposits on the same day. The bankers use statistical approach based on probability to estimate the number of depositors and their claims for a certain day.

(iii) Insurance

The knowledge of statistics is of great use in insurance business. The amount of premium for the insurance policies is based on life tables.

(iv) Sciences

In all sciences i.e., Physics, Chemistry, Biometry, Biology and Psychology, a lot of data are to be analysed. The analysis of data is possible only by an expert in statistics.

1.10. COLLECTION OF DATA

The first step in any investigation is collection of data. The data may be collected for the whole population or for a sample only. It is mostly collected on sample basis. Collection of data is not an easy task. For this job, the staff is trained and work is assigned to them.

1.10.1. PRIMARY DATA

The data are called primary if it is collected originally by some individual or agency and are original in character. The data is called primary if no statistical treatment is applied on the data. There are different methods of collecting the primary data.

1.10.2. SECONDARY DATA

The data are called secondary when the primary data have been treated by some statistical tool. If the primary data have been treated statistically it is no more primary. It becomes secondary data.

1.10.3. COLLECTION OF PRIMARY DATA

Primary data are collected by the following method.

(i) Direct Personal Observations

In this method the enumerator goes to the respondents and interviews them. The information supplied by the informants is recorded by the enumerator. The information collected by this method is considered accurate and complete. The quality of the information (data) depends upon the two persons i.e., the respondent and the enumerator. This method of collecting data is very costly because a team of enumerators is required to get the observations. It is suitable for small local inquiries. It is not suitable if the area of inquiry is wide. Lot of funds and human resources are required for an extensive field of inquiry.

(ii) Indirect Personal Investigation

This method of collecting primary data is adopted when the respondents do not supply the correct information. They do not cooperate with the enumerators. If people are asked to supply information about their income, their houses and other holdings, they are quite likely to give wrong information. In this method some agency other than the respondent is selected to supply the data. For example if data is required from shop keepers, their trade unions can be consulted to supply the data about their members. Indirect questions can also be asked from the respondent so that he speaks out the correct figures. The respondent can be asked about the size of his house, his facilities, and other services like gas, telephone, and electricity etc. By indirect questions the respondent can be forced to give the correct answer.

(iii) Questionnaire Method

Questionnaire is a list of questions relating to the field of inquiry. This list of questions is sent by mail to the respondents with a request that they send it back after duly filling the entries. This is a cheap method but this method is not successful in areas where people are not literate and cooperative. It is possible that the people do not return the questionnaire because they are afraid of giving information in writing. The framing of questionnaire is quite difficult. Only an expert can frame a good questionnaire. A questionnaire should have the following characteristics.

(i) Its language should be simple.

(ii) Questions should have an objective answer like 'yes' or 'no'.

(iii) Number of questions should not be very large.

This method is very popular and is being used in big cities where most of the people are literate.

(iv) Collection through Enumerators

This method is used to overcome the difficulties of the illiterate people in filling the questionnaire. The trained enumerators go to the informants and they assist them in filling the questionnaire correctly. This is a very costly inquiry and can be carried out by the Govt. only.

(v) Collection through Local Sources

These are the reports of the local representatives sent to the investigator. The data is not formally collected by local representatives. They only send the reports. In our country the *Patwari* sends regular reports to the Govt. regarding the crop conditions in our country. These reports are only the estimates based on the judgements.

1.10.4. COLLECTION OF SECONDARY DATA

The secondary data are not collected from individual units but are taken from different sources called secondary sources. Some of these sources are as follows:

(i) Official Statistics

All the government departments maintain their data and publish it annually as official statistics. The important publications are those of the Statistics Division and Bureaus of Statistics in different provinces.

(ii) Semi Official

The data available with PIA, railways, State Bank of Pakistan, district councils etc. are called semi official.

(iii) Secondary data can also be obtained from magazines, news papers and journals of the universities

1.11 ACCURACY

It is impossible to study statistics without an amount of arithmetic work. Although many statistical results are not required to a high degree of accuracy, there is often a risk of considerable loss of accuracy during a calculation. As a working rule two more significant figures should be used in a calculation than are required in the final answer. This means that answers derived using four figure logarithm tables should only be quoted to two significant figures.

1.12 SIGNIFICANT FIGURES

By significant figures we mean those figures in a number which give its information other than its magnitude. The following numbers are all given to four significant figures :

658.7	61.39	832900	0.001592

When a final zero is shown in a decimal fraction, such as, 0.05290 , it must be taken as a significant figure. Significant figures of some numbers are given in the following table .

Number	26.2	0.262	0.00262	2.00	170	6200	4.5800
Significant Figures	3	3	3	3	2	2	5
Number	0.003700	0.0037	0.95	1.00580	378	100.00	0.09001
Significant Figures	4	2	2	6	3	5	4

1.13 ROUNDING OFF FIGURES

When it is required to express figures to a smaller number of significant figures than are available they must be rounded off. The figure following the last significant figure is examined. If it is less than5 then the last significant figure stands as it is, e.g., 523 becomes 520. If it is greater than 5 or 5 followed by any further figures, it is increased by 1, e.g., 523.6 becomes 524 and 523.502 becomes 524. When this figure is exactly 5, there is a problem. Any rule such as " round 5's down" , or "round 5's up" must introduce bias, so the following rule is preferred. Round off the 5 so that the last significant figures become even, e.g., 1975 becomes 1980 and 1965 becomes 1960. This rule should produce as many roundings up as roundings down and so avoid bias.

Example 1.1

Round the following numbers off to three significant figures.

(a) 26243 (b) 2624 (c) 2626 (d) 26.24 (e) 0.2624 (f) 0.002626

(g) 0.005901 (h) 0.005905 (i) 0.3875 (j) 4.315 (k) 0.004500

Solution:

(a) 26200 (b) 2620 (c) 2630 (d) 26.2 (e) 0.262 (f) 0.00263

(g) 0.00590 (h) 0.00590 (i) 0.388 (j) 4.32 (k) 0.00450

Example 1.2

Round the following numbers off to three decimal places.

(a) 2.6666 (b) 7.6665 (c) 9.6755 (d) 10.67 (e) 267

(f) 0.000717 (g) 0.4295 (h) 0.97045 (i) 125.9995 (j) 43.8705

Solution:

(a) 2.667 (b) 7.666 (c) 9.676 (d) 10.670 (e) 267.000

(f) 0.001 (g) 0.430 (h) 0.970 (i) 126.000 (j) 43.870

1.14 MATHEMATICAL NOTATION

In addition to mathematical notations in everyday use, certain other less-common notations will be used; a short description of these is given for reference.

1.14.1 MULTIPLICATION

A cross '×' as well as a stop '.' will be used to indicate that two or more items must be multiplied together ; also where no confusion arises the dot may be omitted.

For example, $X . \log X = X \times \log X = X \log X$

$$a . b = a \times b = ab$$

1.14.2 INEQUALITIES

The inequality signs are used with the following meanings :

$X > Y$	X is greater than Y
$X \geq Y$	X is greater than or equal to Y
$X < Y$	X is less than Y
$X \leq Y$	X is less than or equal to Y.

The sense of the inequality may be remembered by noting that the open end of the sign is placed towards the larger item. A common use of the inequality signs is $15 < X < 20$, meaning that X has a value between 15 and 20, or $15 \leq X \leq 20$, where the values 15 and 20 are also included.

1.14.3 APPROXIMATIONS

$X \simeq Y$, X is approximately equal to Y. This will be used when approximations are used in calculations where less accuracy is required than has been kept in the calculation. It implies that the values of X and Y can be regarded as equal on the particular occasion being considered, although they may be distinguished on other occasions.

1.14.4 LIMITS

An arrow will indicate that a quantity tends to a certain value, and under the described conditions approaches closer and closer to it. For example, $X \to 5$, X tends to 5, implies that X becomes closer to 5, but not that X ever equals 5. We can imagine that $(5 - X)$ is less than any real amount, however small it may be. In particular $X \to 0$, X tends to zero, implies that X becomes very small; $X \to \infty$, X tends to infinity, implies that X becomes very large. Note that $X \to 0$ does not imply that X ever becomes zero, and the size of X implied by $X \to \infty$ will depend on the circumstances.

1.14.5 MODULUS

The sign ' | | ' implies to take the numerical value of the quantity inside the sign and make it positive. For example,

$|5| = 5, \quad |-5| = 5, \quad |15-10| = 5, \quad |10-15| = 5.$

1.14.6 FACTORIALS

If n is a positive whole number, the product $1 \times 2 \times 3 \times \cdots \times n$ is called ' factorial n' and is given by the symbol n!. So

$n! = 1 \times 2 \times 3 \times \cdots \times n$

$3! = 1 \times 2 \times 3 = 6 \qquad 5! = 1 \times 2 \times 3 \times 4 \times 5 = 120.$

$8! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 = 40320.$

Many hand calculators have the facility for calculating factorials. Look for a button marked n! or X!. Zero is the only other number, a part from the positive whole numbers, for which there is a definition of factorial. $0! = 1$. But $(-3)!$ is not defined and so is meaningless.

1.14.7 SUMMATION (Σ)

We will often require the sum of a number of similar quantities; for example, $X_1 + X_2 + X_3 + \cdots + X_n$ where n quantities of the type X are added together. This operation of summation is symbolised by the use of the Greek capital letter Σ. The full notation for the above sum would be

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \ldots + X_n$$

We can see that the symbol Σ (the upper-case version of the Greek letter sigma) stands for the operation of summing. In future, it will help you to remember what Σ means in statistics if you remember " capital sigma means sum". If it is clear how many values are to be summed we simply write ΣX instead of $\sum_{i=1}^{n} X_i$. So Σx means " sum the X values". Here is some other useful notation:

ΣX^2	means square each value of X, and then sum.
$(\Sigma X)^2$	means sum the values of X and then square the total.
$\Sigma (X-3)$	means subtract 3 from each value of X, and then sum.
ΣXY	means multiply each value of X and Y and then sum.

Consider the following examples :

(i) $f_1 + f_2 + f_3 + \ldots + f_n = \sum_{i=1}^{n} f_i$

(ii) $X_1^2 + X_2^2 + X_3^2 + \ldots + X_{10}^2 = \sum_{i=1}^{10} X_i^2$

(iii) $(X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3) + \cdots + (X_n + Y_n) = \sum_{i=1}^{n} (X_i + Y_i)$

(iv) $(X_1 + 3)^2 + (X_2 + 3)^2 + (X_3 + 3)^2 + (X_4 + 3)^2 = \sum_{i=1}^{4} (X_i + 3)^2$

(v) $(a_1b_1) + (a_2b_2) + (a_3b_3) + \cdots + (a_Nb_N) = \sum_{i=1}^{N} (a_ib_i)$

(vi) $(X_1 - 1)1^2 + (X_2 - 2)2^2 + (X_3 - 3)3^2 + (X_4 - 4)4^2 + (X_5 - 5)5^2 = \sum_{i=1}^{5} (X_i - i)i^2$

(vii) $aX_1 + a2X_2 + a3X_3 + a4X_4 + a5X_5 = \sum_{i=1}^{5} a\,i\,X_i$

(viii) $[(X_1 - A) + (X_2 - A) + (X_3 - A) + \cdots + (X_n - A)]^2 = \left[\sum_{i=1}^{n} (X_i - A)\right]^2$

(ix) $f_1X_1 + f_2X_2 + f_3X_3 + \cdots + f_nX_n = \sum_{i=1}^{n} f_iX_i$

(x) $a + a + a + \cdots + a(\text{n times}) = na$

(xi) $a_1 + a_2 + a_3 + \cdots + a_n = \sum_{i=1}^{n} a_i$

(xii) $(aX_1 \pm b) + (aX_2 \pm b) + (aX_3 \pm b) + \cdots + (aX_n \pm b) = \sum_{i=1}^{n} (aX_i \pm b)$.

(xiii) $f_1(Y_1 - a)^2 + f_2(Y_2 - a)^2 + f_3(Y_3 - a)^2 + \cdots + f_{10}(Y_{10} - a)^2 = \sum_{i=1}^{10} f_i(Y_i - a)^2$.

Example 1.3

The following corresponding values of X and Y have been obtained :

X	1	3	4	5	7	9	12	14
Y	12	10	6	9	4	0	−1	−5

Calculate (i) ΣX (ii) ΣY (iii) ΣX^2 (iv) ΣY^2 (v) ΣXY

(vi) $(\Sigma X)^2$ (vii) $(\Sigma X)(\Sigma Y)$ (viii) $\Sigma\, |X - Y|$ (ix) ΣXY^2

Solution:

(i) $\Sigma X = 1 + 3 + 4 + 5 + 7 + 9 + 12 + 14 = 55$

(ii) $\Sigma Y = 12 + 10 + 6 + 9 + 4 + 0 - 1 - 5 = 35$

(iii) $\Sigma X^2 = (1)^2 + (3)^2 + (4)^2 + (5)^2 + (7)^2 + (9)^2 + (12)^2 + (14)^2$
$= 1 + 9 + 16 + 25 + 49 + 81 + 144 + 196 = 521$

(iv) $\Sigma Y^2 = (12)^2 + (10)^2 + (6)^2 + (9)^2 + (4)^2 + (0)^2 + (-1)^2 + (-5)^2$
$= 144 + 100 + 36 + 81 + 16 + 0 + 1 + 25 = 403$

(v) $\Sigma XY = 1(12) + 3(10) + 4(6) + 5(9) + 7(4) + 9(0) + 12(-1) + 14(-5)$
$= 12 + 30 + 24 + 45 + 28 + 0 - 12 - 70 = 57$

(vi) $(\Sigma X)^2 = (55)^2 = 3025$

(vii) $(\Sigma X)(\Sigma Y) = (55)(35) = 1925$

(viii) $\Sigma |X - Y| = |1-12| + |3-10| + |4-6| + |5-9| + |7-4| + |9-0|$
$+ |12+1| + |14+5|$
$= 11 + 7 + 2 + 4 + 3 + 9 + 13 + 19 = 68$

(ix) $\Sigma XY^2 = 1(12)^2 + 3(10)^2 + 4(6)^2 + 5(9)^2 + 7(4)^2 + 9(0)^2 + 12(-1)^2 + 14(-5)^2$
$= 144 + 300 + 144 + 405 + 112 + 0 + 12 + 350 = 1467$

26-09-2012

SHORT DEFINITIONS

Observation

Anything that can be measured or observed is called an observation.

Data

Numbers or measurements that are collected as a result of observations. is called

or

The facts and figures that are collected, analyzed and interpreted.

Population

A population is the set of all units of interest in a particular study.

or

The complete set of all objects possessing some common characteristic of interest is called a population.

Sample

A sample is a subset of data selected from the population.

or

A sample is a part of the population that is under study.

Parameter

A parameter is a numerical measurement describing some characteristic of a population.

or

A parameter is a value associated with a population.

Statistic

A statistic is a value computed from a sample.

or

A statistic is a numerical measurement describing some characteristic of a sample.

Definition of Statistics

Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

or

A body of methods dealing with the collection, description, analysis and interpretation of information that can be given in numerical form.

Characteristics of Statistics

Statistics are aggregate of facts, affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.

Objective of Statistics

The objective of statistics is to make inferences about a population based upon information contained in a sample.

Variable

A variable is a phenomenon that may vary from one individual or object to another.

or

Any characteristic of a person, group, or environment that can vary or differ.

Constant

A quantity which is fixed is called a constant.

or

A value that is unchangeable is called a constant.

Quantitative Data

Quantitative data are observations measured on a numerical scale.

or

Quantitative data are always numeric and indicate how much or how many of something.

Qualitative Data

Qualitative data are observations that are nonnumerical.

or

Data that provide labels or names for a characteristic of an element.

Discrete Data

Data whose possible values are countable is called discrete data.

Continuous Data

Data whose possible values are uncountable and which may assume any value in an interval, is called continuous data.

Quantitative Variable

When the variable to be studied can be reported numerically, the variable is called a quantitative variable.

Qualitative Variable

When the characteristic being studied cannot be recorded in numerical form, it is called a qualitative variable.

Discrete Variable

A discrete variable is one that can assume only certain values within an interval.

or

A discrete variable is limited to some certain numerical values.

Continuous Variable

A continuous variable can take on all values within a specified range.

or

A variable that can assume an unlimited number of intermediate values between a given range.

Independent Variable

A variable is called an independent variable if it is not influenced by another variable.

or

A variable that provides the basis for estimation, is called independent variable.

Dependent Variable

The variable that is being predicted or estimated, is called dependent variable.

or

A variable whose values are thought to be a function of the values of independent variable, is called a dependent variable.

Descriptive Statistics

Descriptive statistics attempt to present data in numerical, graphical or tabular form in order to convey information clearly.

or

Descriptive statistics consists of methods for organizing and summarizing information in a clear and effective way.

Inferential Statistics

Inferential statistics consists of methods of drawing conclusions about a population based on information obtained from a sample.

or

The branch of statistics concerned with using sample data to make an inference about a population is called inferential statistics.

Primary Data

Primary data are those which are collected for the first time and are thus original in character.

or

Data that are collected by any body for some specific purpose and use, are called primary data.

Secondary Data

Secondary data are those which have already been collected by some other person and which have passed through the statistical process at least once.

or

Data that are collected and compiled by any other source, is called secondary data.

MULTIPLE - CHOICE QUESTIONS

1. The science of collecting, organizing, presenting, analyzing and interpreting data to assist in making more effective decisions is called:
   - (a) statistic
   - (b) parameter
   - (c) population
   - (d) statistics
2. Methods of organizing, summarizing, and presenting data in an informative way is called:
   - (a) descriptive statistics
   - (b) inferential statistics
   - (c) theoretical statistics
   - (d) applied statistics

3. The methods used to determine something about a population on the basis of a sample is called:
   (a) inferential statistics (b) descriptive statistics
   (c) applied statistics (d) theoretical statistics
4. When the characteristic being studied is nonnumeric, it is called a:
   (a) quantitative variable (b) qualitative variable
   (c) discrete variable (d) continuous variable
5. When the variable studied can be reported numerically, the variable is called a:
   (a) quantitative variable (b) qualitative variable
   (c) Independent variable (d) dependent variable
6. A specific characteristic of a population is called:
   (a) statistic (b) parameter
   (c) variable (d) sample
7. A specific characteristic of a sample is called:
   (a) variable (b) constant
   (c) parameter (d) statistic
8. A set of all units of interest in a study is called:
   (a) sample (b) population
   (c) parameter (d) statistic
9. A part of the population selected for study is called a:
   (a) variable (b) data
   (c) sample (d) parameter
10. Listing of the data in order of numerical magnitude are called:
    (a) raw data (b) arrayed data
    (c) discrete data (d) continuous data
11. Listing of the data in the form in which these are collected are known as:
    (a) secondary data (b) raw data
    (c) arrayed data (d) qualitative data
12. Data that are collected by any body for some specific purpose and use are called:
    (a) qualitative data (b) primary data
    (c) secondary data (d) continuous data
13. The data which have under gone any treatment previously is called:
    (a) primary data (b) secondary data
    (c) symmetric data (d) skewed data
14. The data obtained by conducting a survey is called:
    (a) primary data (b) secondary data
    (c) continuous data (d) qualitative data
15. The data collected from published reports is known as:
    (a) discrete data (b) arrayed data
    (c) secondary data (d) primary data

16. A survey in which information is collected from each and every individual of the population is known as:
(a) sample survey (b) pilot survey
(c) biased survey (d) census survey

17. Data used by an agency which originally collected them are:
(a) primary data (b) raw data
(c) secondary data (d) grouped data

18. Registration is the source of:
(a) primary data (b) secondary data
(c) qualitative data (d) continuous data

19. Data in the population census reports are:
(a) ungrouped data (b) secondary data
(c) primary data (d) arrayed data

20. Issuing a national identity card is an example of:
(a) sampling (b) statistic
(c) census (d) registration

21. A variable that assumes only some selected values in a range is called:
(a) continuous variable (b) quantitative variable
(c) discrete variable (d) qualitative variable

22. A variable that assumes any value within a range is called:
(a) discrete variable (b) continuous variable
(c) independent variable (d) dependent variable

23. A variable that provides the basis for estimation is called:
(a) dependent variable (b) independent variable
(c) continuous variable (d) qualitative variable

24. The variable that is being predicted or estimated is called:
(a) dependent variable (b) independent variable
(c) discrete variable (d) continuous variable

25. Monthly rainfall in a city during the last ten years is an example of a:
(a) discrete variable (b) continuous variable
(c) qualitative variable (d) independent variable

26. The proportion of females in a sample of 50 accounts officers is an example of a:
(a) parameter (b) statistic
(c) array (d) variable

27. Number of family members in different families in a town is an example of a:
(a) discrete variable (b) continuous variable
(c) dependent variable (d) qualitative variable

28. Colors of flowers is an example of:
(a) quantitative variable (b) qualitative variable
(c) skewed variable (d) symmetric variable

29. If each measurement in a data set falls into one and only one of a set of categories, the data set is called:
(a) quantitative (b) qualitative
(c) continuous (d) constant

30. Any phenomenon which is not measurable is called:
(a) variable (b) constant
(c) parameter (d) attribute

31. A constant can assume values:
(a) zero (b) one
(c) fixed (d) not fixed

32. A value which does not change from one individual to another indiviaual is called:
(a) variable (b) statistic
(c) constant (d) array

33. In the plural sense, statistics means:
(a) numerical data (b) methods
(c) population data (d) sample data

34. In the singular sense, statistics means:
(a) methods (b) numerical data
(c) sample data (d) population data

35. Weight of earth is a:
(a) discrete variable. (b) qualitative variable.
(c) continuous variable. (d) difficult to tell.

36. Weight of students in a class marks is a:
(a) discrete data. (b) continuous data.
(c) qualitative data. (d) constant data.

37. Life of a T.V. tube is a:
(a) discrete variable (b) continuous variable
(c) qualitative variable (d) constant

38. Questionnaire method is used in collecting:
(a) primary data. (b) secondary data.
(c) published data. (d) true data.

39. Census returns are:
(a) primary data. (b) secondary data.
(c) qualitative data. (d) true data.

40. Students divided into different groups according to their intelligence and gender will generate:
(a) quantitative data. (b) qualitative data.
(c) continuous data. (d) constant

41. Statistics are:
(a) aggregate of facts and figures. (b) always true.
(c) always continuous. (d) always qualitative

42. Statistical results are:
 (a) randomly true (b) always true
 (c) not true (d) true on average
43. Statistics does not study:
 (a) constant (b) statistic
 (c) parameter (d) individual
44. A statistical population may consist of:
 (a) finite number of values (b) infinite number of values
 (c) either of (a) and (b) (d) none of (a) and (b)
45. The only continuous variable here is:
 (a) rain fall on different days in a city.
 (b) number of customers entering a store on different days.
 (c) number of flights landing on an airport on different days.
 (d) none of them
46. Example of descriptive statistics is:
 (a) 70 % people in Pakistan live in rural areas.
 (b) 50 % people are likely to vote in the national election.
 (c) 20 % of the bulbs produced in a factory will be defective.
 (d) difficult to tell.
47. Example of inferential statistics is:
 (a) percentage of smokers in Pakistan.
 (b) percentage of skilled workers in a factory.
 (c) estimate of increase in prices in the next year.
 (d) none of the above
48. Statistics are always:
 (a) exact. (b) estimated values.
 (c) constant. (d) population values.
49. Statistics must be:
 (a) comparable. (b) not comparable.
 (c) discrete in nature. (d) qualitative in nature.
50. Given 6 quantities, X_1 through X_6, the correct notation for adding quantities 3 through 6 is:
 (a) $\sum_{i=6}^{3} X_i$ (b) $\sum_{i=1}^{6} X_i$
 (c) $\sum_{i=2}^{N} X_i$ (d) $\sum_{i=3}^{6} X_i$
51. Given: $X_1 = 12, X_2 = 19, X_3 = 10, X_4 = 7$, $\sum_{i=2}^{4} X_i$ equals:
 (a) 36 (b) 48
 (c) 41 (d) 29

52. The symbolic notation $\sum_{i=2}^{n} Y_i$ tells us to:

(a) add all quantities from Y_1 through Y_n

(b) add all quantities from $Y = 2$ through Y_n

(c) add all quantities from $Y = 2$ through $Y = n$

(d) add all quantities from Y_2 through Y_n

53. $\sum_{i=1}^{n} (X_i - A)$ equals:

(a) $\sum_{i=1}^{n} X_i(-A)$ (b) $\sum_{i=1}^{n} X_i - nA$

(c) $nX_i - nA$ (d) $\sum_{i=1}^{n} X_i - A$

54. The figure 22.25 rounded to one decimal place is:

(a) 22.3 (b) 22.1

(c) 22.2 (d) 22

55. The figure 22.15 rounded to one decimal place is:

(a) 22.2 (b) 22.1

(c) 22 (d) 22.3

56. The figure 22.26 rounded to one decimal place is:

(a) 22.2 (b) 22.3

(c) 22.1 (d) 22

57. The figure 22.24 rounded to one decimal place is:

(a) 22.2 (b) 22.3

(c) 22.1 (d) 22

58. How many methods are used for the collection of data?

(a) 4 (b) 3 (c) 2 (d) 1

Answers

1.	(d)	2.	(a)	3.	(a)	4.	(b)	5.	(a)	6.	(b)	7.	(d)	8.	(b)
9.	(c)	10.	(b)	11.	(b)	12.	(b)	13.	(b)	14.	(a)	15.	(c)	16.	(d)
17.	(a)	18.	(b)	19.	(c)	20.	(d)	21.	(c)	22.	(b)	23.	(b)	24.	(a)
25.	(b)	26.	(b)	27.	(a)	28.	(b)	29.	(b)	30.	(d)	31.	(c)	32.	(c)
33.	(a)	34.	(a)	35.	(c)	36.	(b)	37.	(b)	38.	(a)	39.	(a)	40.	(b)
41	(a)	42	(d)	43	(d)	44	(c)	45	(a)	46	(a)	47	(c)	48	(b)
49	(a)	50	(d)	51	(a)	52	(d)	53	(b)	54	(c)	55	(a)	56	(b)
57	(a)	58	(c)												

SHORT QUESTIONS

Q.1 What is meant by a population?

Q.2 Define population and write down any six examples of population.

Q.3 What is meant by a sample?

Q.4 Differentiate between population and sample.

Q.5 Define parameter and statistic.

Q.6 Differentiate between parameter and statistic.

Q.7 Discuss the meaning of statistics in the singular and the plural sense.

Q.8 Discuss the meaning and scope of statistics.

Q.9 Write down the main characteristics of statistics.

Q.10 Define statistics.

Q.11 Write down the definitions of statistics.

Q.12 Define statistics in plural sense.

Q.13 Distinguish between "statistics" and "statistic".

Q.14 Differentiate between variable and constant.

Q.15 Distinguish between discrete variable and continuous variable.

Q.16 Explain the quantitative and qualitative data.

Q.17 Define qualitative variable, giving examples.

Q.18 Differentiate between quantitative data and qualitative data with examples.

Q.19 Name the branches of statistics.

Q.20 Differentiate between descriptive and inferential statistics.

Q.21 Explain the functions of statistics.

Q.22 Explain the importance of statistics in different fields.

Q.23 Name the sources of primary data.

Q.24 Discuss the questionnaire method used in the collection of primary data.

Q.25 Define secondary data.

Q.26 Write down the different sources of secondary data..

Q.27 What is the difference between primary and secondary data?

Q.28 Distinguish between primary and secondary data. Describe some methods of collection of primary data.

Q.29 Write down the different types of investigations.

Q.30 Write down the different kinds of statistical investigations.

Q.31 Given: $X_1 = 3, X_2 = 1, X_3 = 14, X_4 = 2, X_5 = -10, X_6 = 4$. Find $\sum_{i=3}^{4}(X_i + b)$, when $b = 1$.

Ans: 18

Q.32 Given: $X_1 = 2, X_2 = 8, X_3 = -6, X_4 = 1, X_5 = 0$. Find $\sum_{i=1}^{5}(X_i - a)$, when $a = 2$.

Ans: -5

Q.33 Given: $X_1 = 3$, $X_2 = 1$, $X_3 = 14$, $X_4 = 2$, $X_5 = -10$, $X_6 = 4$. Find $\sum_{i=2}^{5}(X_i + a)$, when $a = 2$.

Ans: 15

Q.34 Given: $X_2 = 8$, $X_3 = -6$, $X_4 = 1$. Find $\sum_{i=2}^{4}(X_i - a)$, when $a = 0$.

Ans: 3

Q.35. If $X_4 = 1$, $X_5 = 0$, find $\sum_{i=4}^{5}(X_i - a)$, when $a = 3$.

Ans: -5

Q.36 Given $\sum_{i=1}^{5} X_i^2 = 120$ and $\sum_{i=1}^{5} X_i = 5$. Find $\sum_{i=1}^{5}(X_i^2 + 2X_i + 1)$.

Ans: 135

Q.37 Given $Y_i = a + bX_i$, $\sum_{i=1}^{5} X_i = 100$, $a = 15$ and $b = 2$. Find $\sum_{i=1}^{5} Y_i$.

Ans: 275

Q.38 Given $X_i = a + bY_i$, $\sum_{i=1}^{4} Y_i = 20$, $a = 5$ and $b = -0.5$. Find $\sum_{i=1}^{4} X_i$.

Ans: 10

Q.39 Expand the following summation signs: (a) $\sum_{i=1}^{3}(Y_i^2 + i)$ (b) $\sum_{i=5}^{8}(Y_i + \mu)$

Ans. (a) $Y_1^2 + Y_2^2 + Y_3^2 + 6$ (b) $Y_5 + Y_6 + Y_7 + Y_8 + 4\mu$

Q.40 Given $\sum f = 10$, $\sum fX = 20$ and $\sum fX^2 = 100$. Evaluate $\sum f(X - 2)^2$.

Ans. 60

EXERCISES

Q.1 Determine whether each of the following variables is quantitative or qualitative.

(i) Amount of money spent on books.
(ii) Mode of payment.
(iii) Gender
(iv) Number of local calls made per month
(v) Colour of eyes
(vi) Number of textbooks purchased
(vii) The age of customers in a shop
(viii) The classification of all employees in a college as principal, professors or clerks.
(ix) Length of telephone calls recorded at a switchboard.
(x) Number of family members

Ans: (i) Quantitative (ii) Qualitative (iii) Qualitative (iv) Quantitative
(v) Qualitative (vi) Quantitative (vii) Quantitative (viii) Qualitative
(ix) Quantitative (x) Quantitative

Q.2 Which of the following represents discrete data and which represents continuous data.

(i) Number of houses in different cities.
(ii) The yearly income of shopkeepers.
(iii) The number of flowers on plants.
(iv) The life time of electric bulbs.
(v) The length of rolls of cloth.
(vi) The weights of students in a college.
(vii) Temperature of a patient at different intervals of time.
(viii) Monthly rainfall in a city during the last ten years.
(ix) Total number of shares sold on different days in the stock market.
(x) Speed of an automobile.

Ans: (i) Discrete (ii) Continuous (iii) Discrete (iv) Continuous
(v) Continuous (vi) Continuous (vii) Continuous (viii) Continuous
(ix) Discrete (x) Continuous

Q.3 For the following statements, decide whether the variable is continuous or discrete.

(i) The number of hours for which a light bulb being functional.
(ii) The number of current account balances checked by a firm of auditors each year.
(iii) The number of rooms-with-bathroom in three-star Pakistani hotels.
(iv) The number of failures per 100 hours of operation of a large computer system.

(v) The number of hours lost per 100 hours due to failures of a large computer system.
(vi) The number of cars manufactured by a company each month.
(vii) The annual rainfall in Islamic countries in 1994.
(viii) The number of earthquakes per year in Japan in the period 1940 to 1994.
(ix) The output of *Dhodak* oil rigs in 1990.
(x) The number of times rats turn right in 10 encounters with a T-junction in a maze.
(xi) Number of road accidents on the G.T. road between Lahore and Rawalpindi.
(xii) The weight of rice in kg. produced per acre.
(xiii) The reaction time of rats to a stimulus.
(xiv) The number of errors per page of a balance sheet.
(xv) The yield of tomatoes per plant in a greenhouse.
(xvi) The amount of milk produced by a cow during different months.
(xvii) The number of T.V. licences issued by any branch of a bank each year.
(xviii) The percentage of hydrogen content of gases collected from samples near a volcanic eruption.
(xix) A finite number of values or a countable number of values.

Ans: (i) Continuous (ii) Discrete (iii) Discrete (iv) Discrete
(v) Discrete (vi) Discrete (vii) Continuous (viii) Discrete
(ix) Discrete (x) Discrete (xi) Discrete (xii) Continuous
(xiii) Continuous (xiv) Discrete (xv) Continuous (xvi) Continuous
(xvii) Discrete (xviii) Continuous. (xix) Discrete

Q.4 Identify the following quantitative variables as discrete or continuous.
(i) The number of customers arriving between 10 and 11 A.M at bank.
(ii) Area of an office.
(iii) The inventory of cars for sale at an automobile dealer.
(iv) Distance required for stopping an automobile travelling at 80 kilometers per hour.
(v) Price of a share in Karachi Stock Exchange.

Ans: (i) Discrete (ii) Continuous (iii) Discrete
(iv) Continuous (v) Continuous

Q.5 Write out each of the following summations:

(i) $\sum_{k=4}^{8} X_k$ (ii) $\sum_{i=1}^{4} (Y_i - 3)^2$ (iii) $\sum_{j=1}^{3} (X_j - A)$

(iv) $\sum_{i=3}^{5} X_i^3 f_i$ (v) $\sum_{i=1}^{N} c$ (vi) $\sum_{j=1}^{3} (X_j + Y_j - Z_j)^2$

(vii) $\sum_{k=1}^{4} Y_k^2$ (viii) $\sum_{j=1}^{3} (3X_j - 2)^2$ (ix) $\sum_{i=1}^{n} f_i (X_i - \bar{X})^2$

Ans:

(i) $X_4 + X_5 + X_6 + X_7 + X_8$

(ii) $(Y_1 - 3)^2 + (Y_2 - 3)^2 + (Y_3 - 3)^2 + (Y_4 - 3)^2$

(iii) $(X_1 - A) + (X_2 - A) + (X_3 - A) = X_1 + X_2 + X_3 - 3A$

(iv) $X_3^3 f_3 + X_4^3 f_4 + X_5^3 f_5$

(v) $c + c + c + \ldots + c$ (N times) $= Nc$

(vi) $(X_1 + Y_1 - Z_1)^2 + (X_2 + Y_2 - Z_2)^2 + (X_3 + Y_3 - Z_3)^2$

(vii) $Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2$

(viii) $(3X_1 - 2)^2 + (3X_2 - 2)^2 + (3X_3 - 2)^2$

(ix) $f_1 (X_1 - \bar{X})^2 + f_2 (X_2 - \bar{X})^2 + f_3 (X_3 - \bar{X})^2 + \cdots + f_n (X_n - \bar{X})^2$

Q.6 Use the summation symbol to represent the following expressions:

(i) $X_3 + X_4 + X_5 + \cdots + X_{20}$

(ii) $cX_1 + cX_2 + c X_3 + \cdots + cX_N$

(iii) $f_1 X_1^2 D_1^2 + f_2 X_2^2 D_2^2 + f_3 X_3^2 D_3^2 + \cdots + f_{10} X_{10}^2 D_{10}^2$

(iv) $f_1 (X_1 - a)^2 + f_2 (X_2 - a)^2 + f_3 (X_3 - a)^2$

(v) $f_1(u_1, v_1) + f_2(u_2, v_2) + f_3(u_3, v_3) + \cdots + f_k(u_k, v_k)$

(vi) $a_1(X_1 + 1) + a_2(X_2 + 2) + a_3(X_3 + 3) + \cdots + a_{10}(x_{10} + 10)$

Ans:

(i) $\sum_{i=3}^{20} X_i$ (ii) $\sum_{i=1}^{N} cX_i$ (iii) $\sum_{i=1}^{10} f_i X_i^2 D_i^2$

(iv) $\sum_{i=1}^{3} f_i (X_i - a)^2$ (v) $\sum_{i=1}^{k} f_i (u_i, v_i)$ (vi) $\sum_{i=1}^{10} a_i (X_i + i)$

Q.7 Given $a_1 = 3$, $a_2 = 5$, $a_3 = 8$ and $a_4 = 2$, evaluate

(i) Σa (ii) $\Sigma(6 - a)$ (iii) $\Sigma(4a - 2)$

(iv) Σa^2 (v) $\Sigma(a - 2)^2$ (vi) $\Sigma 2(a - 3)$

Ans:

(i) 18 (ii) 6 (iii) 64 (iv) 102 (v) 46 (vi) 12

Q.8 Find the value for each of the following expressions:

(i) $\sum_{i=1}^{6} 3i$ (ii) $\sum_{j=1}^{5} (j^2 + 3)$ (iii) $\sum_{k=1}^{4} (3k - 2)$

(iv) $\sum_{t=2}^{5} (t^2 + t)$ (v) $\sum_{\ell=2}^{3} (\ell^2 + 3\ell - 2)$ (vi) $\sum_{i=4}^{6} (25 - i^2)$

Ans:

(i) 63 (ii) 70 (iii) 22 (iv) 68 (v) 24 (vi) − 2

Q.9 Given that $\sum_{i=1}^{6} X_i = 14$ and $\sum_{i=1}^{6} X_i^2 = 44$, find

(i) $\sum_{i=1}^{6} (X_i - 1)$ (ii) $\sum_{i=1}^{6} (3X_i + 2)$ (iii) $\sum_{i=1}^{6} (X_i + 4)^2$

Ans:

(i) 8 (ii) 54 (iii) 252

Q.10 If $\sum_{i=1}^{5} X_i = -24$ and $\sum_{i=1}^{5} X_i^2 = 124$, Calculate

(i) $\sum_{i=1}^{5} (2X_i + 10)$ (ii) $\sum_{i=1}^{5} X_i (X_i - 3)$

(iii) $\sum_{i=1}^{5} (X_i - 4)^2$ (iv) $\sum_{i=1}^{5} (X_i + 5)(X_i - 5)$

Ans: (i) 2 (ii) 196 (iii) 396 (iv) − 1

Q.11 Given $\sum_{i=1}^{4} X_i = 20$, $\sum_{i=1}^{4} Y_i = -10$ and $\sum_{i=1}^{4} X_i Y_i = 15$. Find

(i) $\sum_{i=1}^{4} (4X_i + 3Y_i)$ (ii) $\sum_{i=1}^{4} (X_i - 5)(3Y_i + 6)$

(iii) $\sum_{i=1}^{4} X_i (Y_i - 2)$ (iv) $\sum_{i=1}^{4} (X_i + 2)(Y_i - 2)$

Ans:

(i) 50 (ii) 195 (iii) − 25 (iv) − 61

Q.12 Given that Y_t is defined as IBM's sales in billions of dollars in year t. The sales of the IBM corporation in 1966 – 76 are shown below:

Year (t)	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976
Sales (Y_t)	4.2	5.3	6.9	7.2	7.5	8.3	9.5	11.0	12.7	14.4	16.3

Evaluate each of the following expressions:

(i) $\sum_{t=1966}^{1971} Y_t$ (ii) $\sum_{t=1967}^{1970} (Y_t - 3)$ (iii) $\sum_{t=1968}^{1970} (Y_t - 2)^2$ (iv) $\sum_{t=1969}^{1973} 2Y_t^2$

Ans:

(i) 39.4 (ii) 14.9 (iii) 81.3 (iv) 776.46

Q.13 Suppose that we have two series, X and Y, which are paired as follows:

X	1	3	5	7	9	11
Y	10	8	6	4	2	1

Compute: (i) $(\Sigma X)(\Sigma Y)$ (ii) $\Sigma X^2 - (\Sigma X)^2$ (iii) $\Sigma X^2 Y - \Sigma X^2$

(iv) $\Sigma(2X + 3Y)$ (v) $\Sigma(X + Y)(X - Y)$ (vi) $\Sigma X^2(Y + 5)$

Ans:

(i) 1116 (ii) − 1010 (iii) 425 (iv) 165 (v) 65 (vi) 2141

Q.14 Given:

$X_1 = 5$	$X_2 = 7$	$X_3 = 10$	$X_4 = 12$	$X_5 = 15$
$Y_1 = 8$	$Y_2 = 10$	$Y_3 = 13$	$Y_4 = 18$	$Y_5 = 20$

Calculate: (i) ΣX^2 (ii) ΣXY (iii) $\Sigma(X - 7)^2$

(iv) $\Sigma X(Y - 6)$ (v) $\Sigma 3Y^2$ (vi) $\Sigma(X - 5)(Y - 10)$

(vii) $\Sigma(X + Y)(X - Y)$ (viii) $\Sigma 2XY^2$

Ans:

(i) 543 (ii) 756 (iii) 102 (iv) 462

(v) 3171 (vi) 171 (vii) − 514 (viii) 25196

Chapter

2

PRESENTATION OF DATA

2.1. INTRODUCTION

The number of observations is usually very large in the collected data. It is difficult to work on the observations which are called the *raw data*. Before we apply any statistical tool on the data the observations are put into some condensed form so that the statistical work becomes simple. This is called presentation of data. Presentation of data includes (i) Classification (ii) Tabulation (iii) Graphic Presentation.

2.1.1. CLASSIFICATION

The huge amount of data containing individual observations is divided into some classes. Each class contains a set of homogeneous observations. The process of arranging observations into homogeneous groups or classes is called classification. According to L.R. Connor "classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities".

The classification of the data mainly depends upon the nature, scope and purpose of the statistical inquiry. Some characteristics of a good classification are:

(i) Classification should be unambiguous

Different classes should be defined clearly. All the classes should be mutually exclusive and there should be no confusion about them. If the industrial workers are divided into two classes i.e. skilled and unskilled, it should be quite clear as to who is a skilled worker and who is unskilled.

(ii) Classification should be stable

Suppose we have to carry out a survey about smokers in our country. Who is a smoker? It must be defined clearly. Some people are casual smokers and some are regular. Definition of a smoker should not vary from place to place or from time to time. Thus classification of data should be stable so that the data collected from different places on different occasions are comparable.

(iii) Classification should not be rigid

Classification should be able to accommodate the new observations or to drop the observations which are no more required.

2.1.2. TYPES OF CLASSIFICATION

Classification depends upon the characteristics of the statistical data. There are two characteristics namely descriptive and numerical. The examples of

descriptive characteristics are friendship, gender, intelligence and poverty etc. Descriptive characteristics cannot be measured. When the data are classified on the basis of *descriptive statistics,* it is known as classification according to *attributes* or classification according to *qualitative variable.* When the data are classified on the basis of numerical characteristics, it is called classification according to *class intervals* or classification according to *quantitative variable.*

2.2. CLASSIFICATION ACCORDING TO ATTRIBUTES

There may be one, two or more attributes and the observations may be put into different classes depending upon the number of attributes.

2.2.1. SIMPLE CLASSIFICATION

There may be only a single attribute (characteristic) and the data may be divided into different classes on the basis of this attribute. Suppose we are interested to study the population for literates. Some people are literate and some are illiterate. Thus the attribute has two levels (sub-classes). This is called twofold division or *dichotomy* (cutting into two). The attribute may have three levels for example (i) highly educated (ii) average education (iii) not educated. When an attribute has three levels, it is called 3-fold division or *trichotomy*. More than three levels are also possible. For example different income levels of the people may be (i) very rich (ii) rich (iii) average income (iv) low income. The *attribute* of income has four levels (sub-classes). When an attribute has more than 3 levels it is called *manifold-division.* Thus the number of classes based on a single attribute may be two or more than two. As long as there is one attribute, the classification is called *one-way* or *simple.*

2.2.2. TWO-WAY CLASSIFICATION

When the data are divided into different classes on the basis of two attributes, it is called two - way classification. In two-way classification there are at least 4 classes. Suppose one attribute is income with two levels-rich and poor and the other is education with two levels-literate and illiterate, then the data are divided into $2 \times 2 = 4$ classes given below:

(i) rich and literate
(ii) rich and illiterate
(iii) poor and literate
(iv) poor and illiterate

If one attribute has 3 levels and the other has 2 levels, then there are $3 \times 2 = 6$ classes of the data. If the attribute of income has 3 levels-rich, average income and poor and the other attribute education has 2 levels-literate and illiterate, then the data are divided into 6 classes as below:

(i) rich and literate
(ii) rich and illiterate
(iii) average income and literate
(iv) average income and illiterate
(v) poor and literate
(vi) poor and illiterate.

If both attributes have 3 levels, then the data are divided into $3 \times 3 = 9$ classes. As long as there are two attributes, the classification is *two-way*. When their are more than two attributes the classification is called *manifold*.

2.3. QUANTITATIVE DATA

When the observations are generated by counting or by measurement, the data obtained are called quantitative. There are three different methods of presenting the quantitative data.

(i) Series of individual observations. (ii) Discrete frequency distribution.

(iii) Frequency distribution (Grouped Data) or Classification according to class-intervals.

2.3.1. SERIES OF INDIVIDUAL OBSERVATIONS

When the data are collected, these are in the form of individual observations. These are also called *raw data*. When raw numerical data are arranged in ascending or descending order of magnitude, it is called an *array or arrayed data*. An array of observations is called *series of individual observations*. Suppose we have asked 10 students about their marks in a test in English. Their responses are recorded below which are called *raw data*.

R. No. of students	1	2	3	4	5	6	7	8	9	10
Marks in English	30	40	30	40	50	40	50	50	40	40

Raw data when arranged in ascending order become arrayed data or series of individual observations. The array is 30, 30, 40, 40, 40, 40, 40, 50, 50, 50.

2.3.2. DISCRETE FREQUENCY DISTRIBUTION

When the variable is discrete, the observations obtained are called discrete data. Discrete data is generated by counting, therefore each and every observation is exact. When an observation is repeated, it is counted. The 'number' for which the observation is repeated is called 'frequency' of that observation. The symbol 'f' is used for frequency. The symbol X is commonly used for observations. Suppose we have raw data as below which are the numbers of students in ten classes of a school.

30	32	30	40	42	40	32	40	32	40

There are 2 sections of 30 students each, 3 sections have 32 students, 4 sections have 40 students and 1 section has 42 students. This information can be collected in the form of table below called discrete frequency distribution.

Discrete Frequency Distribution of Number of Students

Number of students X	Number of sections f
30	2
32	3
40	4
42	1

This method of condensation of observations into discrete frequency distribution is for discrete data but it can also be used for continuous data as well. In actual practice this method is not used when the number of observations is very large. For very large number of observations, the observations are put into different groups. The grouping can be done for discrete as well as continuous data.

The symbol 'X' is used for the variable which is recorded initially in the form of observations and the symbol 'f' is used for the number of times each observation is repeated.

Example 2.1.

The number of persons in 20 families in a small village are:

2	1	3	2	4	2	4	5	4	5	3	4	5	5	4	5	4	5	4	5

Put the observations in discrete frequency distribution form.

Solution:

Number of persons X	Number of families f
1	1
2	3
3	2
4	7
5	7
Total	20

2.4. FREQUENCY DISTRIBUTION (GROUPED DATA)

The observations on a continuous variable are not exact. These are rather the approximate measurements. We know from the theory of continuous variables that an observation (measurement) does not repeat itself as far as theory is concerned. In actual practice, the observations are repeated. These measurements are put into different classes, each class accommodates a certain number of observations called the frequency (f) of the class. The condensation of data into suitable number of classes, each class containing the observations which lie within the class, is called a *frequency distribution or grouped data.* The grouped data are also called *classification according to class intervals.* This is very popular way of presenting the statistical observations. The observations get condensed into a form convenient for further calculations.

In actual practice the *grouped data* is used for continuous as well as discrete data. For illustration of various terms used in grouped data, we take the weights of 40 students in kilograms (recorded to the nearest kilogram) and put them in grouped data as shown in Table 2.1. The weights are:

56	55	59	58	59	63	63	61	62	61	62	62	65	64	64	65	66	67	69	64
63	61	69	61	58	59	65	64	55	62	63	64	59	62	56	62	64	67	66	65

Table 2.1. Entry Table

Weight (kilograms)	Observations	No. of students (f)
55 – 57	56, 55, 55, 56	4
58 – 60	59, 58, 59, 58, 59, 59	6
61 – 63	63, 63, 62, 61, 62, 62, 61 62, 63, 62, 61, 62, 61, 63	14
64 – 66	65, 64, 64, 65, 66, 64 65, 64, 64, 65, 66, 64	12
67 – 69	67, 69, 69, 67	4

The various terms which are to be explained in connection with the grouped data are

(i) number of classes
(ii) class limits and class boundaries
(iii) size of the interval (class-interval)
(iv) class-marks or mid points
(v) frequency

2.4.1. NUMBER OF CLASSES

There is no hard and fast rule about the number of classes. It is decided on the basis of the given data. Clearly, if the observations are spread over a long range of the variable, we shall make large number of classes. Keeping in view the relative importance of the observations, the observations of large magnitude and small magnitude are not put in the same class. If we take a very small number of classes, we shall loose much of the information contained in the raw data. In the data about heights of students, if the height 60" is combined with the height 72", the calculations on the grouped data will differ to a great extent from the calculations on the original observations. If we take very large number of classes, it will spoil the purpose of classification. Thus we have to make a wise decision about the number of classes. The given observations, their nature, their size and their number give us some hint to decide the number of classes. Roughly, it can be suggested that the number of classes should be between 8 to 15. More than 15 classes may not increase the accuracy in calculations and less than 8 classes will mean that we are combining the large observations with the small observations with the result that their originality will be lost.

In Table 2.1, there are only 5 classes which are 55 – 57, 58 – 60, 61 – 63, 64 – 66, and 67 – 69.

The students are advised to make more than 5 classes when they are given some data for classification.

The starting class is 55 – 57. This class will accommodate the observations equal to 55, 56, and 57. The classes like 55 – 57, 58 – 60, ... are suitable for discrete set of data. But it will not be wrong to use this type of classes for continuous data.

For continuous data, the classes are taken as 55 – 57, 57 – 59, 59 – 61, ... Now 57 is present in the starting class as well as in the second class. When this is the case, the observation 57 is to be recorded in the class 57 – 59 and not in the class 55 – 57. Thus if the end point of a class and the starting point of the next class are equal, the 'next class' will absorb the observations equal to these end points.

The word *inclusive classification* is used for the classes like 55 – 57, 58 – 60, ..., 67 – 69. Both end points of the class are included in the respective class.

For the classes like 55 – 57, 57 – 59, ..., the word *exclusive classification* is used. It is exclusive because the observation 57 is excluded from the class 55 – 57. A class may be "less than 10" or "more than 90". Such classes are called open ended classes. One end of such a class is not specified. Such classes should be avoided because there are certain important calculations (for example the mean, the geometric mean, the harmonic mean, the mean deviation and standard deviation) which cannot be made when the classified data contain some open end classes.

Some authors use the words 'class' and 'class interval' interchangeably. We have not used the word 'class interval' for a 'class'.

2.4.2 CLASS LIMITS AND CLASS BOUNDARIES

The class 55 – 57 starts with 55 and ends at 57. The starting point 55 is called lower class limit and the end point 57 is called upper class limit. The class limits should be taken such that it is clear as to which observation should be included in that class. In continuous data when observations are recorded in whole numbers, the observation 55, in fact, is a value between 54.5 and 55.5. An observation which is 54.6 is recorded as 55. Thus 55 starts from 54.5 which is called the lower class boundary of the group 55 – 57. The upper limit of the group 55 – 57 is 57. The observation 57 covers all observations on a continuous variable up to 57.5. Thus 57.5 is the actual upper boundary of the class 55 – 57. The class 55 – 57 includes all those observations which fall in the expanded interval 54.5 – 57.5. The values 54.5 and 57.5 are the lower and upper class boundaries of the class 55 – 57. A class boundary can be calculated by finding the mean of the upper limit of a class and the lower limit of the next higher class. If the original observations are recorded as whole numbers, the class boundaries are given up to one decimal place. It means no observation will be equal to any class boundary. Thus there will be no confusion whether or not an observation should be included in a class. If the actual observations are measured up to one decimal place, the class boundaries will be up to two decimal places and so on. The data of Table 2.1. are reproduced here in Table 2.2. with class boundaries calculated.

Table 2.2. Class boundaries

Weight (kilogram)	Class boundaries	No. of students f
55 – 57	54.5 – 57.5	4
58 – 60	57.5 – 60.5	6
61 – 63	60.5 – 63.5	14
64 – 66	63.5 – 66.5	12
67 – 69	66.5 – 69.5	4

Suppose we have classes as 1.0 – 1.4, 1.5 – 1.9, 2.0 – 2.4 ... with some frequencies, the class boundaries are calculated as below:

Classes	Class boundaries	f
1.0 – 1.4	0.95 – 1.45	
1.5 – 1.9	1.45 – 1.95	
2.0 - 2.4	1.95 - 2.45	
⋮	⋮	

The first group has the lower class boundary equal to 0.95 which is the average of 1 and 0.9, where 0.9 is the upper limit of a group '0.5 – 0.9' which has not been written in the above table. Upper class boundary of the first group is 1.45 which is the average of 1.4 and 1.5. The value 1.45 is also the lower class boundary of the second group. Similarly 1.95 is the average of 1.9 and 2 where 2 is the lower limit of the next class.

We have to calculate the class boundaries of the grouped data when we have to

(i) Calculate median, quartiles, deciles and percentiles of the data.

(ii) Calculate mode.

(iii) Make graphs of frequency distributions.

For the classes 5 – 10, 10 – 15, 15 – 20, ..., calculation of class boundaries is not required because the given lower and upper class limits are already the class boundaries. For the open ended class "less than 10", the lower class boundary cannot be calculated. Similarly for the open ended class "more than 90" the upper class boundary cannot be calculated.

2.4.3. SIZE OF THE INTERVAL (CLASS INTERVAL)

The difference between the upper class boundary and the lower class boundary of a class is called the 'interval' or 'class interval' of the class. It is also called the size of the interval, width or length of the interval. In Table 2.2, the class boundaries of the starting group are 54.5 and 57.5. Thus the interval of this class is $57.5 - 54.5 = 3$. Interval for a class can also be calculated by finding the difference between the two successive lower class limits or between two successive upper class limits or between the two successive mid points. The difference will be the interval for the preceding of the two classes. All classes may have the same interval which is called the uniform interval. A uniform interval is helpful in statistical calculations. In some special cases, interval of unequal length can also be used. Uniform interval can be denoted by some symbol 'h' or 'i'.

Warning

Table 2.1. has the classes 55 – 57, 58 – 60, ... This is an inclusive classification. In these classes length of the interval calculated by finding the difference between two class limits is wrong. The difference between 57 and 55 is $57 - 55 = 2$, whereas the interval actually is $57.5 - 54.5 = 3$. Thus an interval is the difference between

two class boundaries and not the difference between two class limits. It is always safe to find the difference between two successive lower class limits (or lower class boundaries) for finding interval of the preceding of the two classes. When the classes are 10 – 15, 15 – 20, we may find the difference between two class limits. Here the difference between 15 and 10 is 5 which is the interval of the class 10 – 15. If possible the size of the interval should be 1, 2, 5, 10 or any multiple of these digits or it should be 0.01, 0.1, 0.2, 0.5. We shall avoid the intervals like 5.1, 2.3, 10.4 etc. We prefer the whole numbers as intervals.

2.4.4. CLASS MARKS OR MIDPOINTS

For a certain class, the class mark or midpoint is calculated by finding the mean of the two class limits or the mean of the two class boundaries. It makes no difference whether we calculate it from the two class limits or the two class boundaries. In Table 2.1, we have the classes with class limits i.e., 55 – 57, 58 – 60, ... Their midpoints are $\frac{55+57}{2} = 56$ and $\frac{58+60}{2} = 59$. In Table 2.2, we have the classes with class boundaries 54.5 – 57.5, 57.5 – 60.5, ... The midpoints are $\frac{54.5+57.5}{2} = 56$ and $\frac{57.5+60.5}{2} = 59$ which are same as calculated from the class limits. The class mark of a class represents all the observations of that class. It is a representative figure of the observations of its class. If it is the average or close to the average of the observations lying in that class, then the calculations on the grouped data will be close to the calculations on the actual raw data.

2.4.5. FREQUENCY

Frequency of a class is the number of observations following in that class. There are two methods of recording the observations against the respective classes. These methods are:

(i) By listing the actual observations (entry table)

(ii) By using a tally-column (tally sheet)

In table 2.1. where the observations have been listed against their respective classes the first method is used. The observations against each class are counted and this count is recorded as frequency (f) of each class in a column having heading as frequency. In this table the frequencies of the classes are 4, 6, 14, 12, 4 under the heading 'number of students'.

In Tally Sheet method, we put a stroke or a mark '✓' or vertical line or bar as 'I' for each observation against the respective class under the column headed Tally Bars. Four marks or strokes like ✓✓✓✓ show that 4 observations have gone to a class. The 5th observation is recorded by crossing these 4 strokes to get ~~✓✓✓✓~~ which stands for a group of 5 observations. When all the observations have been marked, we write the total of the strokes corresponding to each class in the column headed by frequency. Groups of five are convenient for counting. The observations in table 2.1. are presented in table 2.3. with the help of tally sheet. The class boundaries and the midpoints are also given.

Table 2.3. Tally Sheet

Frequency Distribution of Weights of 40 Students

Weights (kilograms)	Class boundaries	Midpoints	Tally Bars	Frequency f
55 – 57	54.5 – 57.5	56	✓✓✓✓	4
58 – 60	57.5 – 60.5	59	~~✓✓✓✓~~ ✓	6
61 – 63	60.5 – 63.5	62	~~✓✓✓✓~~ ~~✓✓✓✓~~ ✓✓✓✓	14
64 – 66	63.5 – 66.5	65	~~✓✓✓✓~~ ~~✓✓✓✓~~ ✓✓	12
67 – 69	66.5 – 69.5	68	✓✓✓✓	4

2.4.6. FORMING A FREQUENCY DISTRIBUTION

Grouped data are used for presentation of continuous data but if the discrete data are very large, they are treated as continuous for the sake of convenience of calculations. Hence the discrete data can be grouped in the same way as the continuous data and rest of the calculations are similar for both types of data. For making a frequency distribution we adopt the following procedure:

(i) First of all we find the range of the data. The range is the difference between the maximum and the minimum observation.

(ii) We decide the approximate number of classes into which the data are to be grouped. Suppose the range is 65 and we want to make about 12 classes. The range is divided by the number of classes. We get $65/12 = 5.4$, which is the approximate size of the interval of the classes. We usually avoid intervals like 5.4. We can use interval of 5 or 6. If we take 5 as interval, number of classes will increase. Thus if interval is 5, the number of classes will be $65/5 = 13$. If we take 6 as length of interval, the number of classes will be less than 12 because $65/6 = 10.8$ which means 11 classes. Thus we can take a start with approximate number of classes and then we can adjust the interval so that it is preferably an integer like 1, 2, 3, 4, 5, 10, 20, 25 or 50 or any multiple of these integers. If the interval is less than 1, we shall take interval equal to 0.01, 0.05, 0.1, 0.2, 0.4, 0.5 or 0.8. We can also take a start by taking a reasonable size of the interval and then the number of classes can be determined.

(iii) When the number of classes and the interval has been decided, then we write the class limits. The lowest class usually starts with a number which is a multiple of the interval. The class limits should be written so that the smallest observation is absorbed in the starting class and the largest observation is absorbed in the last group. The open ended classes should be avoided.

(iv) All the observations are put into respective classes. This may be done by using 'tally column' method which is suitable for tabulating the observations into respective classes. The total of the tally bars is calculated to get the frequency against each class. The frequencies of all classes are noted to get grouped data or frequency distribution of the given variable. The total of the frequency column must be equal to the number of observations.

Example 2.2.

Arrange the following marks in a frequency table taking a suitable size of interval by : (i) Entry table (ii) Tally sheet.

13	81	58	81	85	75	61	70	84	81	87	67	65	62	62	61	59	58	57	75
72	84	91	87	76	43	83	40	73	86	73	43	33	76	95	73	65	77	72	72
29	43	85	42	80	75	85	62	57	64	70	95	57	74	50	78	49	55	64	92
73	73	96	69	57	22	78	80	36	70	85	47	69	63	53	91	33	69	30	34

Solution:

Here, highest observation = 96, lowest observation = 13, Range = 96 – 13 = 83. If we take an interval of 10, we shall have 9 classes which is a reasonable number. The minimum observation is 13. Let us take the lowest class as 10 – 20. The frequency distribution is obtained as below:

(i) By Entry Table

Classes	Observations	f
10 – 20	13	1
20 – 30	29, 22	2
30 – 40	33, 36, 33, 30, 34	5
40 – 50	43, 40, 43, 43, 42, 49, 47	7
50 – 60	58, 59, 58, 57, 57, 57, 50, 55, 57, 53	10
60 – 70	61, 67, 65, 62, 62, 61, 65, 62, 64, 64, 69, 69, 63, 69	14
70 – 80	75, 70, 75, 72, 76, 73, 73, 76, 73, 77, 72, 72, 75, 70, 74, 78, 73, 73, 78, 70	20
80 – 90	81, 81, 85, 84, 81, 87, 84, 87, 83, 86, 85, 80, 85, 80, 85	15
90 – 100	91, 95, 95, 92, 96, 91	6
Total	----	80

(ii) By Tally Sheet

Classes	Observations	f
10 – 20	✓	1
20 – 30	✓✓	2
30 – 40	~~✓✓✓✓~~	5
40 – 50	~~✓✓✓✓~~ ✓✓	7
50 – 60	~~✓✓✓✓~~ ~~✓✓✓✓~~	10
60 – 70	~~✓✓✓✓~~ ~~✓✓✓✓~~ ✓✓✓✓	14
70 – 80	~~✓✓✓✓~~ ~~✓✓✓✓~~ ~~✓✓✓✓~~ ~~✓✓✓✓~~	20
80 – 90	~~✓✓✓✓~~ ~~✓✓✓✓~~ ~~✓✓✓✓~~	15
90 – 100	~~✓✓✓✓~~ ✓	6
Total	---	80

2.4.7. RELATIVE FREQUENCY DISTRIBUTION

If frequency of a class is divided by the sum of frequencies we get what is called a *relative frequency*. If we calculate the relative frequencies for all the classes, we get the relative frequency distribution. The total of the relative frequencies is equal to 1. If the frequencies in Table 2.1. are converted into *relative frequencies,* we get relative frequency table or relative frequency distribution. Table 2.4. gives the relative frequency distribution of weights of 40 students.

Table 2.4.

Relative Frequency Distribution of Weights of 40 students

Weights (kilograms)	Frequency	Relative frequencies
55 – 57	4	$\frac{4}{40} = 0.10$
58 – 60	6	$\frac{6}{40} = 0.15$
61 – 63	14	$\frac{14}{40} = 0.35$
64 – 66	12	$\frac{12}{40} = 0.30$
67 – 69	4	$\frac{4}{40} = 0.10$
Total	40	1.00

The relative frequencies are also called proportions and in discussion on probability we shall call them probabilities of the classes. The idea of relative frequencies is helpful in understanding the basic lessons on probability. It is also used in the normal distribution and other probability distributions where the total area under the curve is unity.

2.4.8. PERCENTAGE FREQUENCY DISTRIBUTION

If a relative frequency is multiplied by 100, we get percentage relative frequency. If all the relative frequencies are converted into percentage relative frequencies, we get percentage relative frequency distribution or simply percentage frequency distribution. The relative frequencies of Table 2.4. are converted into percentage relative frequencies which are given in Table 2.5.

Table 2.5.

Percentage Frequency Distribution of Weights of 40 students

Weights (kilograms)	Frequency	Percentage frequencies
55 – 57	4	$\frac{4}{40} \times 100 = 10\ \%$
58 – 60	6	$\frac{6}{40} \times 100 = 15\ \%$
61 – 63	14	$\frac{14}{40} \times 100 = 35\ \%$
64 – 66	12	$\frac{12}{40} \times 100 = 30\ \%$
67 – 69	4	$\frac{4}{40} \times 100 = 10\ \%$
Total	40	100 %

2.4.9. CUMULATIVE FREQUENCY DISTRIBUTION

For cumulative frequency distribution, the class limits are converted into class boundaries. Cumulative frequency of a class is the total of all frequencies up to that class. Thus in Table 2.5. cumulative frequency of the class 57.5 – 60.5 is 4 + 6 = 10 and the cumulative frequency of the class 60.5 – 63.5 is 4 + 6 + 14 = 24. This means that there are 10 observations less than 60.5 and there are 24 observations less than 63.5. These are called 'less than' cumulative frequencies. If we calculate the cumulative frequencies from the bottom, we get what are called "more than" cumulative frequencies. Thus there are 4 observations more than 66.5, there are 4 + 12 = 16 observations more than 63.5 and there are 4 + 12 + 14 = 30 observations more than 60.5. The "less than" and "more than" cumulative frequencies of Table 2.1. are given in Table 2.6(a) and Table 2.6(b) respectively.

Table 2.6(a)

Less than Cumulative frequency distribution of weights of 40 students

Weights	Frequency	Class boundaries	Weights 'less than'	'less than' cumulative frequency
55 – 57	4	54.5 – 57.5	less than 57.5	4
58 – 60	6	57.5 – 60.5	less than 60.5	4 + 6 = 10
61 – 63	14	60.5 – 63.5	less than 63.5	10 + 14 = 24
64 – 66	12	63.5 – 66.5	less than 66.5	24 + 12 = 36
67 – 69	4	66.5 – 69.5	less than 69.5	36 + 4 = 40

Table 2.6(b)

'More than' cumulative frequency distribution of weights of 40 students

Weights	Frequency	Class boundaries	Weights 'more than'	'more than' cumulative frequency
55 – 57	4	54.5 – 57.5	more than 54.5	36 + 4 = 40
58 – 60	6	57.5 – 60.5	more than 57.5	30 + 6 = 36
61 – 63	14	60.5 – 63.5	more than 60.5	16 + 14 = 30
64 – 66	12	63.5 – 66.5	more than 63.5	4 + 12 = 16
67 – 69	4	66.5 – 69.5	more than 66.5	4

When it is not mentioned 'less than' or 'more than' cumulative frequency then it always means the 'less than' cumulative frequency. The cumulative frequency distribution gives us information about the number of observations which are less than or more than some particular point.

The percentage frequency distribution can also be converted into cumulative frequency distribution. It is called cumulative percentage frequency distribution. Percentage frequency distribution of Table 2.5. converted into cumulative percentage frequency distribution is given in Table 2.7.

Table 2.7

Cumulative percentage frequency distribution of weights of 40 students

Weights	Frequency	Class boundaries	Percentage frequency	Weight 'less than'	Cumulative percentage frequency
55 – 57	4	54.5 – 57.5	$\frac{4}{40} \times 100 = 10\ \%$	less than 57.5	10 %
58 – 60	6	57.5 – 60.5	$\frac{6}{40} \times 100 = 15\ \%$	less than 60.5	25 %
61 – 63	14	60.5 – 63.5	$\frac{14}{40} \times 100 = 35\ \%$	less than 63.5	60 %
64 – 66	12	63.5 – 66.5	$\frac{12}{40} \times 100 = 30\ \%$	less than 66.5	90 %
67 – 69	4	66.5 – 69.5	$\frac{4}{40} \times 100 = 10\ \%$	less than 69.5	100 %
Total	40	--	100 %	---	--

It may be mentioned here that the cumulative frequencies of Table 2.6(a) can also be converted into percentage frequencies. Each cumulative frequency of Table 2.6(a) is divided by 40 and multiplied by 100 to get the result given in the last column of Table 2.7.

Thus it makes no difference whether the frequencies are first converted into percentage frequencies and then cumulated (as is done in Table 2.7) or the cumulative frequencies are calculated first and are then converted into percentage frequencies. The students are advised to convert the cumulative frequencies of Table 2.6(a) into cumulative percentage frequencies.

2.4.10. BI-VARIATE FREQUENCY DISTRIBUTION

We have discussed the frequency distributions of a single variable. Such frequency distributions are called univariate which means a single variable. Sometimes we have observations in pairs and each pair is obtained as a result of observations on two variables. For example the marks of mathematics and statistics for a student is a single pair of observations. The marks of another student in two subjects will make another pair of observations. Suppose we have the ages of 65 husbands and their wives given below:-

Ages of husbands	16	20	20	25	18	26	26	30	29	30	26	27	21
Ages of wives	20	17	18	20	16	27	25	25	30	25	25	23	19
Ages of husbands	29	29	25	24	23	26	30	29	21	23	25	28	18
Ages of wives	25	24	20	27	18	23	21	30	19	25	18	25	20
Ages of husbands	18	17	20	25	30	25	25	17	28	29	28	27	20
Ages of wives	20	16	22	16	27	23	25	24	25	27	24	28	25
Ages of husbands	26	25	24	26	27	21	21	22	23	24	27	31	27
Ages of wives	26	25	23	25	27	26	20	20	25	26	30	30,	23
Ages of husbands	28	26	20	22	24	24	20	27	26	29	16	19	26
Ages of wives	29	26	20	22	24	20	23	30	28	30	16	20	18

The ages of husbands and wives are put into 5 groups with classes as 15 – 18, 19 – 22, 23 – 26, 27 – 30 and 31 – 34. The ages of husbands are given in the columns and the ages of wives in rows. An observation (16, 20) means the age of husband is 16 and the age of his wife is 20. This observation is recorded under the first column '15 – 18' and against the second row 19 – 22. At the intersection of 1st column and 2nd row the observation is recorded in a cell by a stroke '1'. All the observations are assigned to respective cells and then we add the strokes in rows and columns to get frequencies in rows and columns. The observations classified according to this method are shown in Table 2.8.

Table 2.8.

Ages of husbands

	Classes	15 – 18	19 – 22	23 – 26	27 – 30	31 – 34	Total
	15 – 18	III	II	IIII			9
Ages	19 – 22	III	~~IIII~~ III	III	I		15
of	23 – 26	I	III	~~IIII~~ ~~IIII~~ IIII	~~IIII~~ IIII		27
Wives	27 – 30			III	~~IIII~~ ~~IIII~~	I	14
	Total	7	13	24	20	1	65

The bivariate frequency distribution obtained from Table 2.8 is given in Table 2.8(a).

Table 2.8(a)

Bivariate frequency distribution of ages of husbands and wives.

Ages of husbands

	Classes	15 – 18	19 – 22	23 – 26	27 – 30	31 – 34	Total
	15 – 18	3	2	4	—	—	9
Ages	19 – 22	3	8	3	1	—	15
of	23 – 26	1	3	14	9	—	27
Wives	27 – 30	—	—	3	10	1	14
	Total	7	13	24	20	1	65

2.5. TABULATION

After collection of raw data the data are divided into different classes on the basis of the characteristics under study. There may be one or more characteristics under study. The classified data, whether it be qualitative or quantitative is arranged in different rows and columns. This orderly arrangement of data in columns and rows is called *Tabulation*. It is the presentation of data in a manner which is suitable for further statistical work and statistical record. For publication, the huge mass of data are presented in the form of suitable *tables*. Tremendous amount of data regarding population of Pakistan are presented in the form of different tables. For such data, it is not possible to do condensation so that these can be presented in a single table. Sometimes a large number of tables are required to highlight different aspects of the data. The size and number of tables also depend upon the volume of data and the number of variables and attributes being studied. Tabulation is a stage where data are ready for reading, for quick understanding, for publication and for further statistical work.

2.5.1. TYPES OF TABULATION

Sometimes the data are stored in the books, magazines and other publications for future reference. These tables are usually large in size and are called *general purpose tables*. Some tables contain the data for analysis and research. These tables are usually small in size and are called *specific purpose tables*.

When the qualitative or quantitative raw data are classified according to one characteristic or one variable, the tabulation of different groups is called single or one-way. The tabulation is two-way or double when there are two characteristics or two variables. Similarly the tabulation may be manifold or complex when the data are divided into different categories on the basis of more than two criteria.

2.5.2. ONE-WAY TABULATION

When the data are tabulated into different categories according to one characteristic, the tabulation is called *simple or one-way*. It is good time to mention a common mistake about one-way tabulation. About one-way tabulation, the impression is that this tabulation has only two groups of data like intelligent and non-intelligent, smokers and non-smokers, literate and illiterate. The data having two groups is though one way tabulation but more categories under the same characteristic are also possible and the table will still be simple or one way. In Table 2.8 the qualitative data are given in one way tabulation. There are 1000 workers working in a factory. They are divided into two groups, skilled and unskilled.

Table 2.8. One-Way Table

No. of skilled workers	No. of unskilled workers	Total
600	400	1000

The workers may be divided into more than two categories. Table 2.9 is another example of one-way table.

Table 2.9. One-Way Table

No. of skilled workers	No. of semi–skilled workers	No. of unskilled workers	Total
600	500	200	1300

In Table 2.10, the quantitative data are given in a *one-way* table.

Table 2.10. One-Way Table

Wages of workers	No. of workers
2000 – 2100	70
2100 – 2200	120
2200 – 2300	250
2300 – 2400	280
2400 – 2500	200
2500 and above	80

Tables 2.8. and 2.9. are about a single characteristic that is skill of the job and table 2.10. is about a single variable 'income'. All these three tables provide answer to a single question. Is the worker skilled? Answer is yes or no. What is the income of the worker? Answer is any one income group. Table 2.10. does not show whether a worker is skilled or not.

2.5.3. TWO-WAY TABULATION

When the data are classified according to two criteria, the tabulation is called *two-way* or *double-tabulation*. Table 2.11. shows two-way table on the data about the workers in a factory.

Table 2.11. Two - Way Table

Wages of workers	No. of skilled workers	No. of unskilled workers	Total
2000 – 2100	40	30	70
2100 – 2200	80	40	120
2200 – 2300	150	100	250
2300 – 2400	150	130	280
2400 – 2500	120	80	200
2500 and over	60	20	80
Total	600	400	1000

Two questions about a worker can be answered from the two-way table. One question about professional skill and the other about the income of the worker.

2.5.4. THREE-WAY TABULATION

When there are three variables or characteristics, the data is put into three-way tabulation Table 2.12 gives the data of factory workers in three-way table according to their wages, their professional training and their religion.

Table 2.12. Three – Way Table

Wages of workers skill wise and religion wise

	Number of skilled workers			Number of unskilled workers			Grand Total
Wages	Muslims	Non-Muslims	Total	Muslims	Non-Muslims	Total	
2000 – 2100	35	5	40	27	3	30	70
2100 – 2200	70	10	80	37	3	40	120
2200 – 2300	130	20	150	90	10	100	250
2300 – 2400	140	10	150	120	10	130	280
2400 – 2500	110	10	120	75	5	80	200
2500 and over	55	5	60	18	2	20	80
Sub-Total	540	60	600	367	33	400	1000
Grand Total	600			400			

A three-way table can provide answers to three different questions regarding the observations with data. For example the three questions may be about income, religion and skill.

2.5.5. MAIN PARTS OF STATISTICAL TABLE

Main parts of a statistical table are the title, the box-head, the stub, the body, prefatory note, footnote and the source note.

Let us sketch here Table 2.13 without figures to explain various sections of the table.

Table 2.13.

Wages of workers ⟵ Title

Prefatory note **Box-head**

Wages	No. of skilled workers			No. of unskilled workers			Grand Total
	Muslims	Non-Muslims	Total	Muslims	Non-Muslims	Total	
Stub→				**Body**			

Foot note.

Source note.

(i) Title

A clear-cut heading of the table is called its title. It is said that the title should answer the questions like *what, where, how classified and when* about the data. Main title should be in capital letters. Different parts of the heading can be separated by commas but no full-stop should be used in the title.

(ii) Box-head

The headings and sub-headings in columns are called column captions. The space where these captions (column headings) are written is called box-head. The captions should be clearly written in prominent letters. First column should contain more important characteristic. Only the first word in each column should be in capital letters.

(iii) Stub

The heading and sub-heading of the rows are called *row captions* and the space where these headings are written is called stub. The stub should contain clear headings.

(iv) Body of data

This is the main part of the table. Below the column headings and against the row marking, we write the figures. It is the main part of the table and contains the information regarding different groups.

(v) Prefatory Note and Footnote

Sometimes the additional notes are given on the table to explain various points which are otherwise not clear. Prefatory note is given below the title to indicate as to which units have been included and which have been ignored. This note is written

below the title and above the box head. Footnote is used to clarify anything in the table. A footnote may be about columns, rows or the body. Footnote is at the end of the table.

(vi) Source Note

The source note is given at the end of the table. It includes the information about compiling agency, publication, date of publication etc.

2.5.6. GENERAL RULES OF TABULATION

There are certain rules which should be followed in tabulation.

(i) A table should be simple, easy to understand. There should be no need to go through footnotes or explanations.

(ii) If the observations are large in number they can be broken into different groups and more than one table can be prepared.

(iii) Proper and clear headings for columns and rows should be used.

(iv) Thick lines should be used to seperate the data under big classes and thin lines to seperate the sub-classes of data.

(v) The units of measurement under each heading and sub-heading must be indicated.

(vi) Totals of columns should be preferably at the bottom of the table and the totals of rows should be at the extreme right side of the table.

It is not possible to make rigid rules for tabulation. In general tables should suit the needs and requirements of an investigation.

2.6. GRAPHIC REPRESENTATION

Statistical tables contain data in the form of figures. But numerical figures are usually not attractive and some people find it difficult to get a clear picture from the numerical data. A more attractive method of presenting the data is to make good looking diagrams and graphs. There are very large number of graphs, pictures and diagrams which are used to represent data. Good looking diagrams and pictures leave an everlasting impression on the mind of the observer. The modern advertising is mainly based on pictures of various types to attract the consumers.

2.6.1. DIAGRAMS

There are different types of diagrams used in representing data. We shall discuss here only the bar diagrams, rectangles and pie diagram.

2.6.2. SIMPLE BAR DIAGRAMS

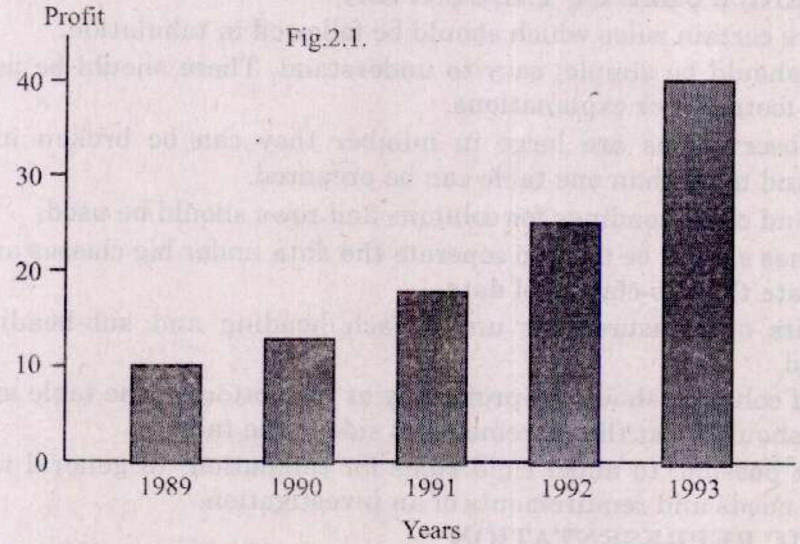
Simple bar diagrams are made to represent geographical, historical, numerical and the qualitative data. The vertical or horizontal bars are made to represent the data when the difference between different quantities is not very large. The different quantities may be arranged in ascending or descending order but the time series data are not arranged. The height of the bars is in proportion to the size of the quantities. All the bars are of uniform width. The space between the bars should not be more than the width of the bars. The bars should be neither very tall nor they should look small statured. Very thick or very thin bars will not look attractive. Vertical bars are used for time series and quantitative data and horizontal bars are used for geographical data.

Example 2.3.

Draw simple bar diagram to represent the profits of a bank for 5 years.

Years	1989	1990	1991	1992	1993
Profit (million Rs.)	10	12	18	25	40

Simple bar diagram showing the profits of a bank for 5 years



2.6.3. MULTIPLE BAR DIAGRAMS

Multiple bar diagrams are made when the data are in the form of different groups and each group contains two or more pieces of information. Within the group, the different quantities are represented by adjacent bars. One set of bars is seperated from another by a suitable space between them. Imports and exports of a country may be shown side by side for different months. Birth and death rates of a country over a number of years may be shown side by side for different years.

Example 2.4.

Draw suitable diagram for the data of birth rates and death rates of following countries.

Country	Birth rate	Death rate
Pakistan	45	15
India	40	16
Britain	10	10
Egypt	35	12
Japan	10	10

The given data are suitable for multiple bar diagrams. Thus the multiple bar diagram is drawn below:

Multiple bar diagram showing birth and death rates in some countries.

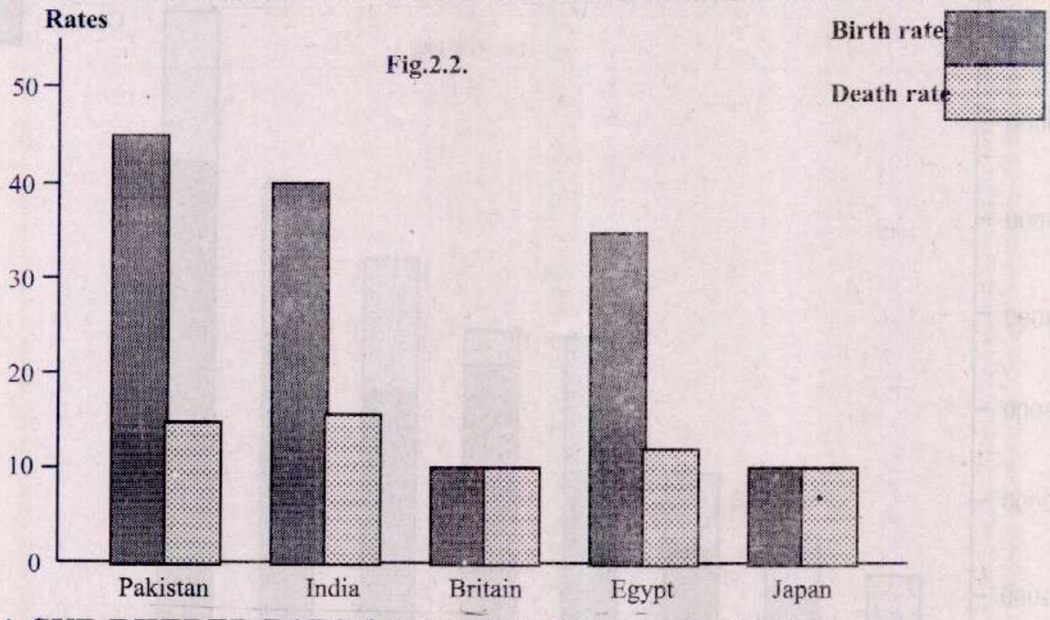


Fig.2.2.

2.6.4. SUB-DIVIDED BARS OR COMPONENT BAR CHART

Component bar charts are made when we are interested to compare the totals of some groups of data and the comparison within the group is also desired. Each bar represents the total of some components and each component is shown as a proportional section of the bar. In this way the quantities within the group and between the groups are compared. The bars are seperated by suitable space between them.

Example 2.5.

Represent the following data about the number of sheep and goats (in hundreds) in various divisions of Punjab during the year 1982 – 83 by component bar chart.

Division	No. of sheep (00)	No. of goats (00)	Total
Bahawalpur	667	1583	2250
D.G. Khan	413	768	1181
Faisalabad	1311	3122	4433
Gujranwla	1495	2265	3760
Lahore	4262	5777	10039
Multan	1434	3778	5212
Rawalpindi	372	3415	3787
Sargodha	448	1048	1496

Source: Punjab Development Statistics, 1984

Component bar chart

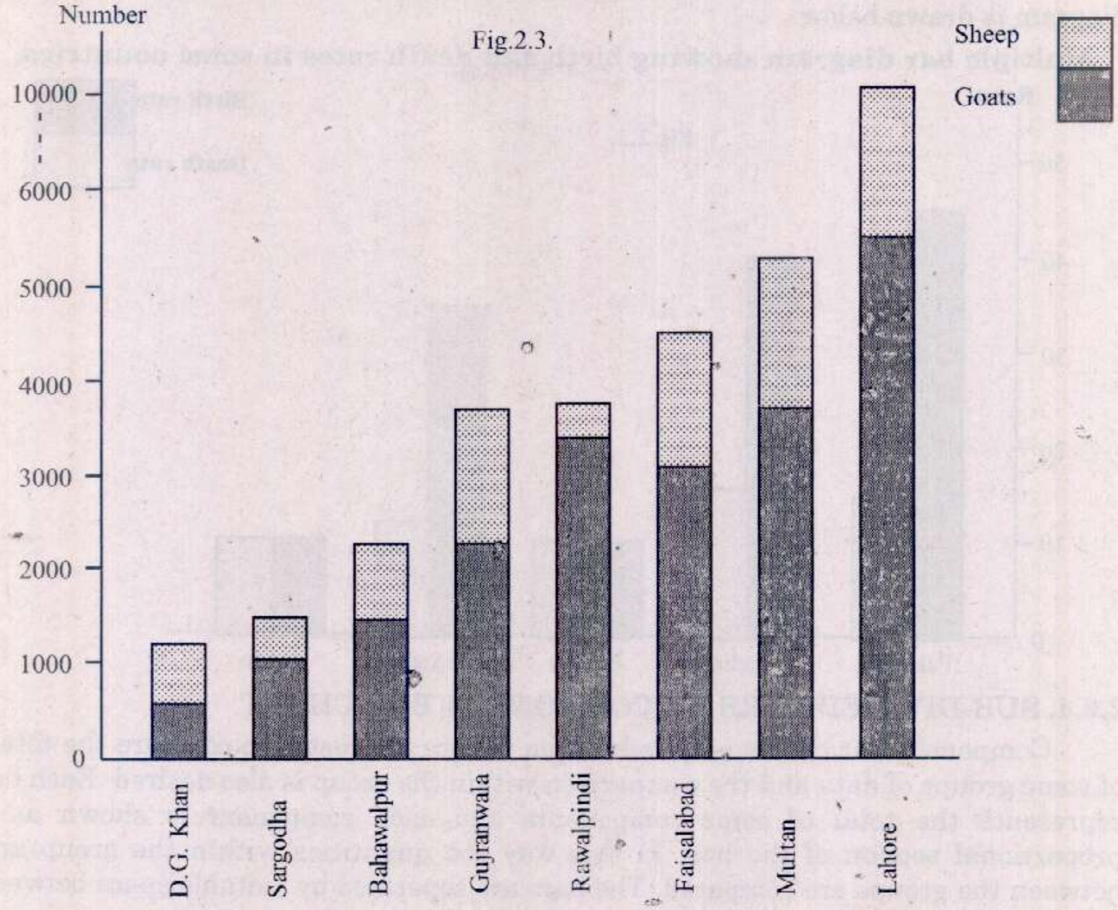


Fig.2.3.

2.6.5. RECTANGLES

The rectangle is four sided geometrical shape with some length and height. The area of the rectangle is given by the product length × height. Rectangles are drawn for those quantities which differ a lot from one another. Rectangles are area diagrams. Rectangle has two dimensions, it is therefore possible to keep one of the dimensions constant for all rectangles and the other in proportion to the area of the rectangle. Suppose a factory has 1000 workers out of which 600 are skilled and 400 are unskilled.

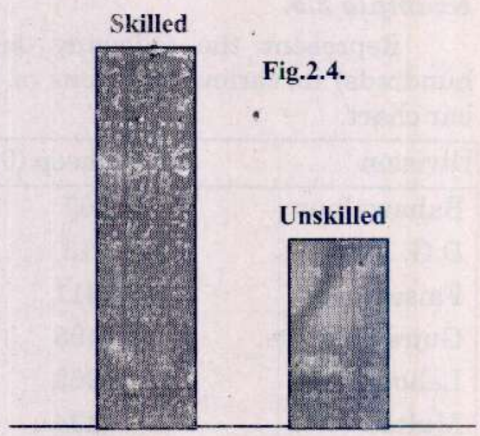
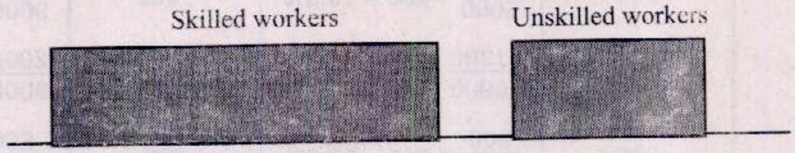


Fig.2.4.

We have to show the skilled and unskilled workers in rectangles. The areas of the two rectangles should be in the proportion 600 : 400 or 3 : 2. The two rectangles with equal base (breadth or width is constant) and the height in proportion to 3 : 2 are

shown in Fig.2.4. These two rectangles look like thick bars. *Thus the rectangles with equal width are nothing but simple thick bars.* We can make rectangles with the same height but length (base) proportional to the ratio 3 : 2. The rectangles with the same height are shown here in Fig.2.5. Their lengths (base) are in the ratio 3 : 2. We can also make rectangles in which both the lengths and heights are different but this type of diagrams are not popular. Rectangles are not very popular diagrams but sub-divided rectangles are very popular

Fig.2.5.



2.6.6. SUB-DIVIDED RECTANGLES

The sub-divided rectangles are drawn to represent two or more groups of data when each group consists of some components and these components are also to be compared. The sub-divided rectangles are usually drawn to compare the budgets of various families. Each component of a group is converted into the percentage of the corresponding total. We draw one rectangle for each total taking equal length (100 units) and breadth is taken proportional to the totals. Each rectangle is divided into different parts which represent different items of expenditure in family budgets.

Example 2.6.

Compare the budgets of family A and B with a suitable diagram.

Items of expenditure	Family A	Family B
Food	2500	4000
Clothing	800	1000
House rent	1200	2000
Educati n	800	600
Miscellaneous	700	1400
Total	6000	9000

Total expenditures are 6000 and 9000 which have the ratio 1 : 1.5. The height of the two rectangles is the same and their breadth is in the ratio 1 : 1.5.

Expenditure on each item is converted into the percentage of its total. The calculations are made here and the sub-divided rectangles are shown in Fig.2.6

Item of expenditure	Family A		Family B	
	Actual expenditure	Percentage expenditure	Actual expenditure	Percentage expenditure
Food	2500	$\frac{2500}{6000} \times 100 = 41.7\%$	4000	$\frac{4000}{9000} \times 100 = 44.4\%$
Clothing	800	$\frac{800}{6000} \times 100 = 13.3\%$	1000	$\frac{1000}{9000} \times 100 = 11.1\%$
House rent	1200	$\frac{1200}{6000} \times 100 = 20\%$	2000	$\frac{2000}{9000} \times 100 = 22.2\%$
Education	800	$\frac{800}{6000} \times 100 = 13.3\%$	600	$\frac{600}{9000} \times 100 = 6.7\%$
Miscellaneous	700	$\frac{700}{6000} \times 100 = 11.7\%$	1400	$\frac{1400}{9000} \times 100 = 15.6\%$
Total	6000	100 %	9000	100 %

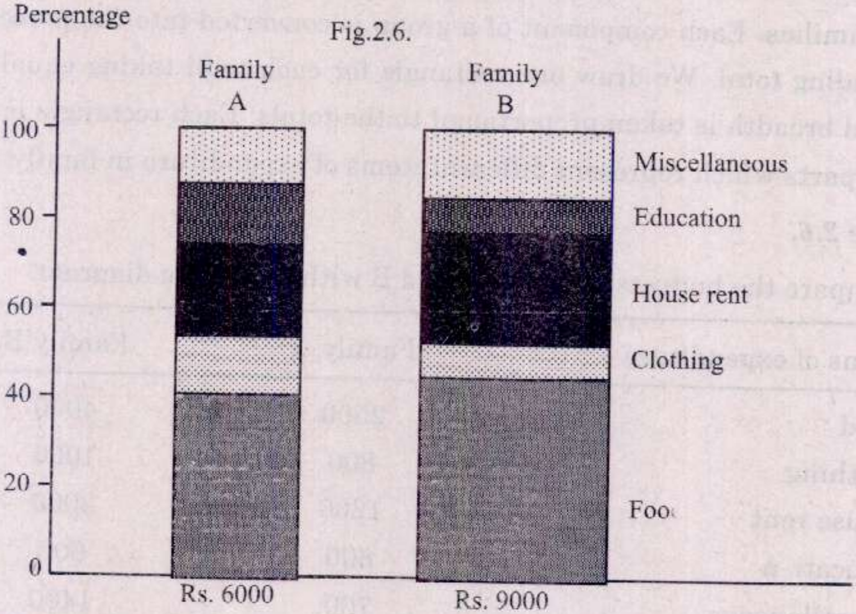


Fig.2.6.

2.6.7. PIE DIAGRAM

Pie diagram is a circular diagram where the whole circle represents a 'total' and the components of the total are represented by sectors of the pie diagram. Pie diagram is also called sector diagram. It is a popular diagram and is drawn when the components are to be shown for comparison. The total angle of the circle is 360° and the total quantity to be represented is taken equal to 360°. The angles for each component are calculated and these angles are made in the circle to show different components.

Example 2.7.

The data on Agricultural Product at current factor cost for Pakistan for the year 1983 – 84 is given below. Make a pie diagram to represent the data.

Sub-sector	Product (million Rs.)
Major crops	46231
Minor crops	14971
Livestock	27096
Fishing	3082
Forestry	457

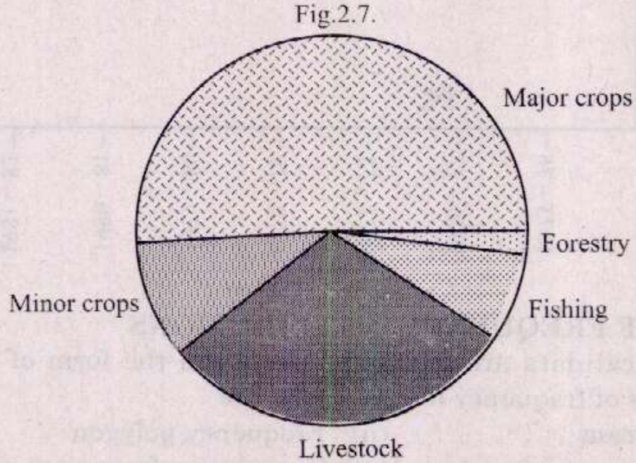
Source: Punjab Development Statistics, 1984.

The necessary calculations to make the Pie diagram are shown below and the diagram is shown in Fig.2.7.

Sub-sectors	Agriculture Product (million Rs.)	Angles of a sub-sectors
Major crops	46231	$\frac{46231}{91837} \times 360 = 181.2$
Minor crops	14971	$\frac{14971}{91837} \times 360 = 58.7$
Livestock	27096	$\frac{27096}{91837} \times 360 = 106.2$
Fishing	3082	$\frac{3082}{91837} \times 360 = 12.1$
Forestry	457	$\frac{457}{91837} \times 360 = 1.8$
Total	91837	360

Pie Diagram showing product of agricultural sub-sectors for 1983 – 84

Fig.2.7.



2.6.8. GRAPH OF TIME SERIES

A graph of time series or historical series is called historigram. The time is taken on X-axis and the variable is taken on Y-axis. The plotted points are joined together to get the graph called historigram.

Example 2.8.

The number of slaughter houses in Punjab for the period 1975 – 76 to 1982 – 83 are given below. Represent the data with a historigram.

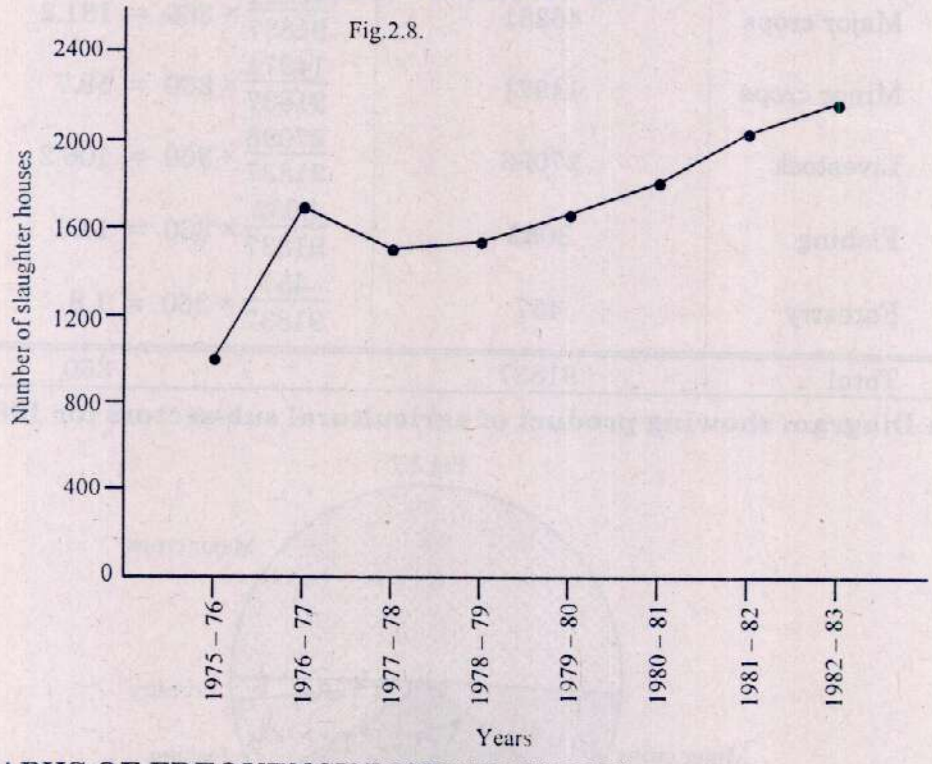
Years	Number of slaughter houses	Years	Number of slaughter houses
1975 – 76	1076	1979 – 80	1641
1976 – 77	1754	1980 – 81	1804
1977 – 78	1546	1981 – 82	2041
1978 – 79	1569	1982 – 83	2172

Source: Punjab Development Statistics, 1984.

The given data is shown in a Historigram in Fig.2.8.

Historigram: Showing number of slaughter houses in Punjab from 1975 – 76 to 1982 – 83

Fig.2.8.



2.7. GRAPHS OF FREQUENCY DISTRIBUTIONS

The statistical data are mostly presented in the form of graphs. Some of the important graphs of frequency distributions are

(i) Histogram (ii) Frequency polygon

(iii) Frequency curve (iv) Cumulative frequency polygon.

2.7.1. HISTOGRAM

Histogram is a graph of the frequency distribution in which classes with class boundaries are taken on X-axis with a suitable breadth of class and adjacent bars are erected to show the frequencies. The height of the bars is in proportion to the size of the frequency. For uniform intervals, we take a suitable breadth for classes. For unequal intervals we have to adjust the frequency. If the interval becomes double, then frequency is divided by 2 so that the area of the bar is in proportion to the areas of other bars. Histogram is a very simple and very important graph of the frequency distribution. This graph makes the base for other graphs. If we take the frequencies on Y-axis, we get frequency histogram, the total area of which is equal to the total frequency. If we take relative frequencies on the Y-axis, the total area of the histogram is unity. If we take the percentage frequencies on Y-axis, we get percentage frequency histogram, the total area of the histogram will be 100.

Example 2.9.

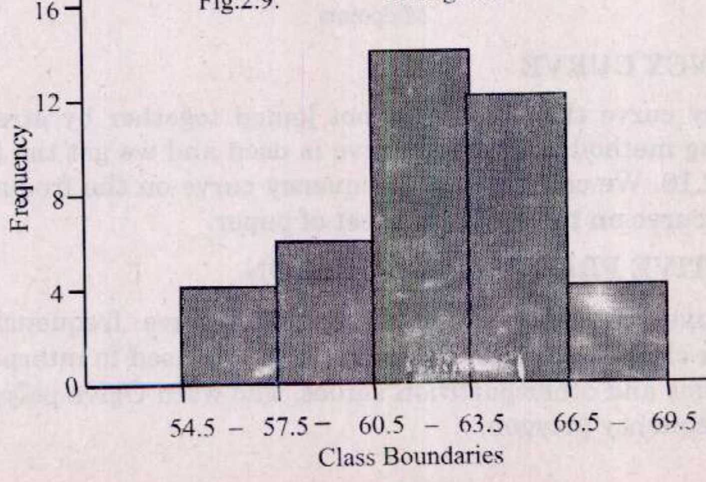
Draw histogram of the following frequency distribution.

Classes	55 – 57	58 – 60	61 – 63	64 – 66	67 – 69
Frequency	4	6	14	12	4

We write the class boundaries in the following table and the Histogram is shown in Fig.2.9.

Groups	Frequency	Class boundaries
55 – 57	4	54.5 – 57.5
58 – 60	6	57.5 – 60.5
61 – 63	14	60.5 – 63.5
64 – 66	12	63.5 – 66.5
67 – 69	4	66.5 – 69.5
Total	40	

Fig.2.9. Histogram



2.7.2. FREQUENCY POLYGON

Frequency polygon is a graph of the frequency distribution in which the frequencies are plotted against the midpoints of the classes. The plotted points are joined together to get the frequency polygon.

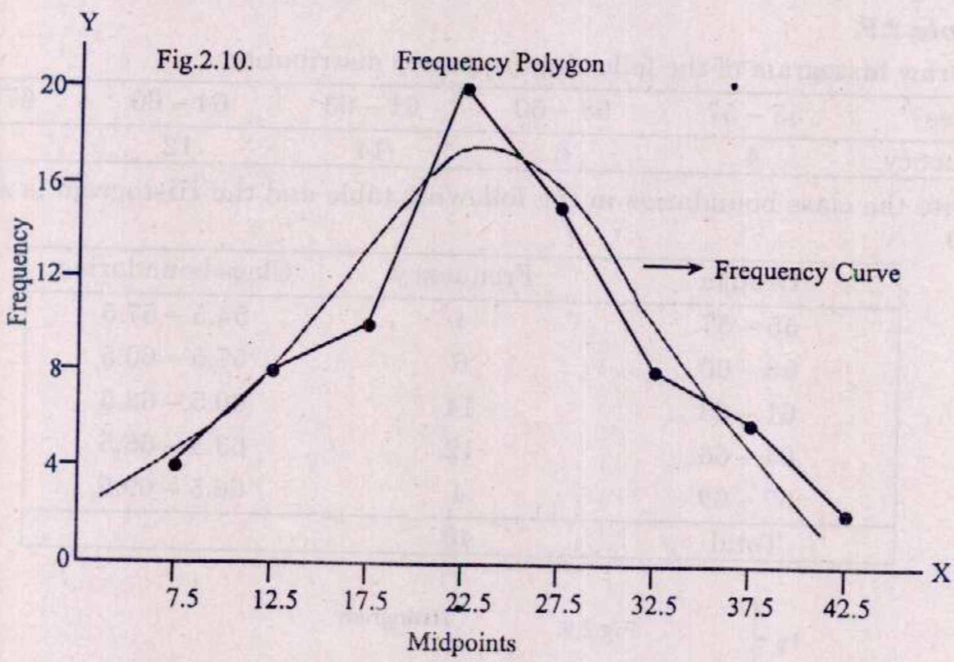
Example 2.10.

Make frequency polygon of the data.

Groups	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45
Frequency	4	8	10	20	15	8	6	2

The midpoints of the classes are taken on X-axis and the frequency is plotted on Y-axis.

The graph is given below in Fig.2.10.



2.7.3. FREQUENCY CURVE

In frequency curve the points are not joined together by straight lines. The free-hand drawing method of drawing curve is used and we get the frequency curve as shown in fig.2.10. We can draw the frequency curve on the frequency polygon or we can draw the curve on the seperate sheet of paper.

2.7.4. CUMULATIVE FREQUENCY POLYGON

In cumulative frequency polygon, the cumulative frequencies are plotted against the upper class boundaries. This graph can be used to interpolate the values of median, quartiles and other partition values. The word Ogive polygon is also used for cumulative frequency polygon.

Example 2.11.

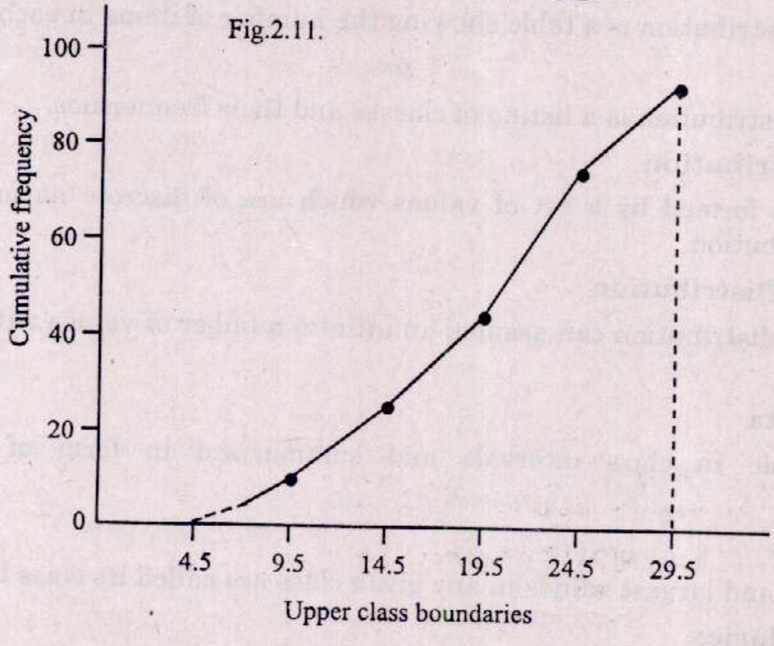
Make cumulative frequency polygon of the given data.

Groups	5 – 9	10 – 14	15 – 19	20 – 24	25 – 29
Frequency	10	15	20	30	18

The necessary calculations are given below:

Groups	Frequency	Class boundaries	Cumulative frequency
5 – 9	10	4.5 – 9.5	10
10 – 14	15	9.5 – 14.5	25
15 – 19	20	14.5 – 19.5	45
20 – 24	30	19.5 – 24.5	75
25 – 29	18	24.5 – 29.5	93

Cumulative Frequency polygon.



SHORT DEFINITIONS

Classification

The process of grouping a large number of individual facts or observations on the basis of similarity among the items is called classification.

or

The process of arranging data in groups or classes according to ~~resemblances and~~ similarities is called classification.

Simple Classification

When data are classified according to a single characteristic, it is called simple classification or one-way classification.

Raw Data

A list of the data in the form in which these are collected, is called raw data.

Array

An array is a listing of the data in order of numerical magnitude.

or

A list of the observations on a variable, either in ascending order or descending order.

Frequency Distribution

A frequency distribution is a table showing the number of items in each class. it is Denoted by f

or

A frequency distribution is a listing of classes and their frequencies.

Discrete Distribution

A distribution formed by a set of values which are of discrete nature is called a discrete distribution.

Continuous Distribution

A continuous distribution can assume an infinite number of values within a specific range.

Grouped Data

Data available in class intervals and summarized in form of a frequency distribution. is Called Grouped Data.

Class Limits

This Upperclass limit and Lower limit is call

The smallest and largest values in any given class are called its class limits.

Class Boundaries

The class boundaries are obtained by increasing the upper class limits and decreasing the lower class limits by the same amount so that there are no gaps between consecutive classes.

Class Width

Lower class

The class width is the difference between two consecutive lower class limits or between the upper and lower class boundaries of any class. (interval)

Class Midpoint

Same of / The half of the ~~difference between~~ the upper and Lower class limit is called Mid point

The point in each class that is halfway between the lower and upper class limits, is called midpoint of that class.

Relative Frequency

The relative frequency for a particular class is equal to the class frequency divided by the total number of observations.

Percentage Frequency

Percentage frequency is computed by dividing the frequency in a category by the total number of observations and multiplying by 100.

Cumulative Frequency

The cumulative frequency is the number of observations less than or equal to a given value of the variable.

or

A cumulative frequency is the sum of the frequencies for several consecutive classes of a frequency distribution.

Bivariate Frequency Distribution or Joint Frequency Distribution

A summary of a bivariate set of data that displays the number of observations of respective joint characteristics of one value taken from each of the variables that define the data set, is called a bivariate frequency distribution.

Tabulation

Tabulation is an orderly arrangement of data in columns and rows.

Pie Chart

Pie chart is a graphical device for presenting qualitative data summaries based on subdivision of a circle into sectors that correspond to the relative frequency for each class.

or

A graph in the shape of a circle that is divided into "slices" corresponding to the categories or classes to be displayed. The size of the slices is proportional to the magnitude of the displayed variable associated with each category or class is called a pie chart.

Histogram

A histogram is a vertical bar chart in which the rectangular bars are constructed at the boundaries of each class.

or

A graphic presentation of data by adjacent rectangles in which each rectangle having area proportional to class frequency is called histogram.

Frequency Polygon

A graphic device for illustrating the frequency distribution in which the frequencies are plotted against class marks and successive points are connected with the help of straight lines.

Frequency Curve

A form of graph, representing a frequency distribution, in which a continuous line is used to indicate the general trend of frequencies.

Cumulative Frequency Polygon or Ogive

The graphical display for a cumulative frequency distribution with upper class boundaries is called a cumulative frequency polygon or ogive.

Cumulative Frequency Curve

A curve that shows the number of cases below the upper true limit of an interval, is called a cumulative frequency curve.

Historigram

A graph of time series or historical series is called historigram.

Skewed Distribution

A distribution that departs from symmetry and tails off at one end is called a skewed distribution.

or

A distribution in which the observations are concentrated at one end of the distribution is called a skewed distribution.

Positively Skewed Distribution

A distribution that has relatively fewer frequencies at the higher end of the horizontal axis, is termed as positively skewed distribution.

Negatively Skewed Distribution

A distribution that has relatively fewer frequencies at the lower end of the horizontal axis.

Normal Curve

A normal curve is a bell shaped frequency curve.

MULTIPLE - CHOICE QUESTIONS

1. When data are classified according to a single characteristic, it is called:
   (a) quantitative classification (b) qualitative classification
   (c) area classification (d) simple classification
2. Classification of data by attributes is called:
   (a) quantitative classification (b) chronological classification
   (c) qualitative classification (d) geographical classification

3. Classification of data according to locations or areas is called:
   (a) qualitative classification (b) quantitative classification
   (c) geographical classification (d) chronological classification
4. Classification is applicable in case of:
   (a) normal characters (b) quantitative characters
   (c) qualitative characters (d) both (b) and (c)
5. In classification, the data are arranged according to:
   (a) similarities (b) differences
   (c) percentages (d) ratios
6. When data are arranged at regular intervals of time, the classification is called:
   (a) qualitative (b) quantitative
   (c) chronological (d) geographical
7. When an attribute has more than three levels it is called:
   (a) manifold-division (b) dichotomy
   (c) one-way (d) bivariate
8. The series

Country	Pakistan	India	Britain	Egypt	Japan
Birth rate	45	40	10	35	10

   is of the type:
   (a) discrete (b) continuous
   (c) individual (d) time series
9. The series

Country	Pakistan	India	Britain	Egypt	Japan
Death rate	15	16	10	12	10

   is of the type:
   (a) inclusive (b) exclusive
   (c) geographical (d) time series
10. In an array, the data are:
   (a) in ascending order (b) in descending order
   (c) either (a) or (b) (d) neither (a) or (b)
11. The number of tally sheet count for each value or a group is called:
   (a) class limit (b) class width
   (c) class boundary (d) frequency
12. The frequency distribution according to individual variate values is called:
   (a) discrete frequency distribution (b) cumulative frequency distribution
   (c) percentage frequency distribution (d) continuous frequency distribution

13. A series arranged according to each and every item is known as:
(a) discrete series (b) continuous series
(c) individual series (d) time series

14. A frequency distribution can be:
(a) qualitative (b) discrete
(c) continuous (d) both (b) and (c)

15. The following frequency distribution

X	5	15	38	47	68
f	2	4	9	3	1

is classified as:
(a) relative frequency distribution (b) continuous distribution
(c) percentage frequency distribution (d) discrete distribution

16. Frequency distribution is often constructed with the help of:
(a) entry table (b) tally sheet
(c) both (a) and (b) (d) neither (a) and (b)

17. The data given as, 3, 5, 15, 35, 70, 84, 96 will be called as:
(a) individual series (b) discrete series
(c) continuous series (d) time series

18. Frequency of a variable is always in:
(a) fraction form (b) percentage form
(c) less than form (d) integer form

19. Data arranged in ascending or descending order of magnitude is called:
(a) ungrouped data (b) grouped data
(c) discrete frequency distribution (d) arrayed data

20. The grouped data are called:
(a) primary data (b) secondary data
(c) raw data (d) difficult to tell

21. A series of data with exclusive classes along with the corresponding frequencies is called:
(a) discrete frequency distribution (b) continuous frequency distribution
(c) percentage frequency distribution (d) cumulative frequency distribution

22. In an exclusive classification, the limits excluded are:
(a) upper limits (b) lower limits
(c) both lower and upper limits (d) either lower or upper limit

23. The series

Weights (pounds)	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40
No. of items	10	15	30	10	5

is categorized as:

(a) continuous series (b) discrete series

(c) time series (d) geometric series

24. The series

Year	2007	2008	2009	2010	2011
Profits (000 Rs.)	7	10	16	18	22

will be called as:

(a) time series (b) discrete series

(c) continuous series (d) individual series

25. The suitable formula for computing the number of classes is:

(a) 3.322 log N (b) 0.322 log N

(c) 1+ 3.322 log N (d) 1 – 3.322 log N

26. The number of classes in a frequency distribution is obtained by dividing the range of variable by the:

(a) total frequency (b) class interval

(c) mid-points (d) relative frequency

27. If the number of workers in a factory is 256, the number of classes will be:

(a) 8 (b) 9 (c) 10 (d) 12

28. The largest and the smallest values of any given class of a frequency distribution are called:

(a) class Intervals (b) class marks

(c) class boundaries (d) class limits

29. If there are no gaps between consecutive classes, the limits are called:

(a) class limits (b) class boundaries

(c) class intervals (d) class marks

30. The extreme values used to describe the different classes in a frequency distribution are called:

(a) class intervals (b) class boundaries

(c) class limits (d) cumulative frequency

31. If in a frequency table, either the lower limit of first class or the upper limit of last class is not a fixed number, then classes are called:

(a) one-way classes (b) two-way classes

(c) discrete classes (d) open-end classes

32. The class boundaries can be taken when the nature of variable is:

(a) discrete (b) continuous

(c) both (a) and (b) (d) qualitative

33. Class boundaries are also called:
(a) mathematical limits (b) arithmetic limits
(c) geometric limits (d) qualitative limits

34. The average of lower and upper class limits is called:
(a) class boundary (b) class frequency
(c) class mark (d) class limit

35. The lower and upper class limits are 20 and 30, the midpoints of the class is:
(a) 20 (b) 25 (c) 30 (d) 50

36. A frequency distribution that contains a class with limits of "10 and under 20" would have a midpoint:
(a) 10 (b) 14.9
(c) 15 (d) 20

37. If the number of workers in a factory is 128 and maximum and minimum hourly wages are 100 and 20 respectively. For the frequency distribution of hourly wages, the class interval is:
(a) 8 (b) 9 (c) 10 (d) 80

38. Width of interval **h** is equal to:
(a) $\dfrac{\text{largest number} - 20}{\text{number of classes}}$ (b) $\dfrac{\text{largest number} + \text{smallest number}}{\text{number of classes}}$
(c) $\dfrac{\text{largest number} - \text{smallest number}}{\text{number of classes}}$ (d) $\dfrac{\text{number of classes}}{\text{range}}$

39. Length of interval is calculated as:
(a) the difference between upper limit and lower limit
(b) the sum of upper limit and lower limit
(c) half of the difference between upper limit and lower limit
(d) half of the sum of upper limit and lower limit

40. The class marks are given below:
10, 12, 14, 16, 18. The first class of the distribution is:
(a) 9 – 12 (b) 10.5 – 12.5
(c) 9 – 11 (d) 10 – 12

41. If the midpoints are 10, 15, 20, 25 and 30. The last class boundary of the distribution is:
(a) 25 – 30 (b) 27.5 – 32.5
(c) 20 – 35 (d) 30 – 35

42. The number of classes depends upon:
(a) class marks (b) frequency
(c) class interval (d) class boundary

43. The class interval is the difference between:
(a) two extreme values
(b) two successive frequencies
(c) two successive upper limits
(d) two largest values

44. When the classes are 40 – 44, 45 – 49, 50 – 54, ⋯, the class interval is:
(a) 4
(b) $\frac{40 + 44}{2}$
(c) 100
(d) 5

45. A grouping of data into mutually exclusive classes showing the number of observations in each class is called:
(a) frequency polygon
(b) relative frequency
(c) frequency distribution
(d) cumulative frequency

46. The following frequency distribution

Classes	Less than2	Less than4	Less than6	Less than8	Less than10
Frequency	2	6	16	19	20

is classified as:
(a) inclusive classification
(b) exclusive classification
(c) discrete classification
(d) cross classification

47. The following frequency distribution

Classes	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	2	4	6	4	2

is classified as:
(a) exclusive classification
(b) inclusive classification
(c) geographical classification
(d) two-way classification

48. The following frequency distribution

Classes	0 – 4	5 – 9	10 – 14	15 – 19	20 – 24
Frequency	2	3	7	5	3

is classified as:
(a) multiple classification
(b) qualitative classification
(c) inclusive classification
(d) exclusive classification

49. The following frequency distribution

Classes	More than4	More than6	More than8	More than10	More than12
Frequency	2	5	10	2	1

is classified as:
(a) geographical classification
(b) chronological classification
(c) inclusive classification
(d) exclusive classification

50. The class frequency divided by the total number of observations is called:

(a) percentage frequency (b) relative frequency
(c) cumulative frequency (d) bivariate frequency

51. The relative frequency multiplied by 100 is called:

(a) percentage frequency (b) cumulative frequency
(c) bivariate frequency (d) simple frequency

52. In a relative frequency distribution, the total of the relative frequencies is:

(a) 100 (b) one
(c) Σf (d) ΣX

53. In a percentage frequency distribution, the total of the percentage frequencies is always equal to:

(a) 1 (b) Σf
(c) 100 % (d) ΣX

54. The cumulative frequency of first group in more than cumulative frequency distribution is always equal to:

(a) 1 (b) 100
(c) Σf (d) ΣX

55. The cumulative frequency of last class in less than cumulative frequency distribution is always equal to:

(a) Σf (b) ΣfX
(c) 1 (d) 100

56. The following frequency distribution

Classes	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50
f	5	15	30	40	48

is classified as:

(a) less than cumulative frequency distribution
(b) more than cumulative frequency distribution
(c) discrete frequency distribution
(d) cumulative percentage frequency distribution

57. The following frequency distribution

Classes	50 – 55	55 – 60	60 – 65	65 – 70	70 – 75
Frequency	40	36	30	16	4

is classified as:

(a) relative frequency distribution
(b) less than cumulative frequency distribution
(c) more than cumulative frequency distribution
(d) bivariate frequency distribution

58. A frequency distribution formed considering two variables at a time is called:
(a) univariate frequency distribution
(b) bivariate frequency distribution
(c) trivariate frequency distribution
(d) bimodal distribution

59. The sum of rows or sum of columns of a bivariate frequency distribution is equal to:
(a) ΣX (b) ΣfX
(c) $\Sigma(f+X)$ (d) Σf

60. The arrangement of data in rows and columns is called:
(a) classification (b) tabulation
(c) frequency distribution (d) cumulative frequency distribution

61. When the qualitative or quantitative raw data are classified according to one characteristic, the tabulation of different groups is called:
(a) dichotomy (b) manifold-division
(c) bivariate (d) one-way

62. A statistical table consists of at least:
(a) two parts (b) three parts
(c) four parts (d) five parts

63. In a statistical table, prefatory note is shown:
(a) below the body (b) box head
(c) foot note (d) below the title

64. A source note in a statistical table is given:
(a) at the end of a table (b) in the beginning of a table
(c) in the middle of a table (d) below the body of a table

65. In a statistical table, column captions are called:
(a) box head (b) stub
(c) body (d) title

66. In a statistical table, row captions are called:
(a) box head (b) stub
(c) body (d) title

67. The headings of the rows of a table are called:
(a) prefatory notes (b) titles
(c) stubs (d) captions

68. The headings of the columns of a table are called:
(a) stubs (b) captions
(c) footnotes (d) source notes

69. The budgets of two families can be compared by:
(a) sub-divided rectangles (b) pie diagram
(c) both (a) and (b) (d) histogram

70. Total angle of the pie-chart is:
(a) 45 (b) 90 (c) 180 (d) 360

71. Diagram are another form of:
(a) classification (b) tabulation
(c) angle (d) percentage

72. In pie diagram, the angle of a sub-sector is obtained as:
(a) $\frac{\text{component part}}{\text{total}} \times 90°$ (b) $\frac{\text{component part}}{\text{total}} \times 180°$
(c) $\frac{\text{component part}}{\text{total}} \times 120°$ (d) $\frac{\text{component part}}{\text{total}} \times 360°$

73. A pie diagram is represented by a:
(a) rectangle (b) circle
(c) triangle (d) square

74. A sector diagram is also called:
(a) bar diagram (b) histogram
(c) historigram (d) pie diagram

75. Which of the followings is not a one-dimensional diagram?
(a) simple bar diagram (b) multiple bar diagram
(c) component bar diagram (d) pie diagram

76. Which of the followings is a two-dimensional diagram?
(a) sub-divided bar (b) percentage component bar chart
(c) sub-divided rectangles (d) multiple bar diagram

77. Pie diagram represents the components of a factor by:
(a) circles (b) sectors
(c) angles (d) percentages

78. The suitable diagram to represent the data relating to the monthly expenditure on different items by a family is:
(a) historigram (b) histogram
(c) multiple bar diagram (d) pie diagram

79. A graph of time series or historical series is called:
(a) histogram (b) historigram
(c) frequency curve (d) frequency polygon

80. The historigram is the graphical presentation of data which are classified:
(a) geographically (b) numerically
(c) qualitatively (d) according to time

81. Historigram and histogram are:
(a) always same (b) not same
(c) off and on same (d) randomly same

82. A distribution in which the observations are concentrated at one end of the distribution is called a:
(a) symmetric distribution (b) normal distribution
(c) skewed distribution (d) uniform distribution

83. For graphic presentation of a frequency distribution, the paper to be used is:
(a) carbon paper (b) ordinary paper
(c) graph paper (d) butter paper

84. Histogram can be drawn only for:
(a) discrete frequency distribution (b) continuous frequency distribution
(c) cumulative frequency distribution (d) relative frequency distribution

85. Histogram is a graph of:
(a) frequency distribution (b) time series
(c) qualitative data (d) ogive

86. Histogram and frequency polygon are two graphical representations of:
(a) frequency distribution (b) class boundaries
(c) class intervals (d) class marks

87. Frequency polygon can be drawn with the help of:
(a) historigram (b) histogram
(c) circle (d) percentage

88. In a cumulative frequency polygon, the cumulative frequency of each class is plotted against:
(a) mid-point (b) lower class boundary
(c) upper class boundary (d) upper class limit

89. The graph of the cumulative frequency distribution is called:
(a) histogram (b) frequency polygon
(c) pictogram (d) ogive

90. When successive mid-points in a histogram are connected by straight lines, the graph is called a:
(a) historigram (b) ogive
(c) frequency curve (d) frequency polygon

91. A frequency polygon is a closed figure which is:
(a) one sided (b) two sided
(c) three sided (d) many sided

92. Ogive curve can be occurred for the distribution of:
(a) less than type (b) more than type
(c) both (a) and (b) (d) neither (a) and (b)

93. The word ogive is also used for:
(a) frequency polygon (b) cumulative frequency polygon
(c) frequency curve (d) histogram
94. Cumulative frequency polygon can be used for the calculation of:
(a) mean (b) median
(c) mode (d) geometric mean

Answers

1.	(d)	2.	(c)	3.	(c)	4.	(d)	5.	(a)	6.	(c)	7.	(a)	8.	(c)
9.	(c)	10.	(c)	11.	(d)	12.	(a)	13.	(c)	14.	(d)	15.	(d)	16.	(c)
17.	(a)	18.	(d)	19.	(d)	20.	(b)	21.	(b)	22.	(d)	23.	(a)	24.	(a)
25.	(c)	26.	(b)	27.	(b)	28.	(d)	29.	(b)	30.	(c)	31.	(d)	32.	(b)
33.	(a)	34.	(c)	35.	(b)	36.	(c)	37.	(c)	38.	(c)	39.	(a)	40.	(c)
41.	(b)	42.	(c)	43.	(c)	44.	(d)	45.	(c)	46.	(b)	47.	(a)	48.	(c)
49.	(d)	50.	(b)	51.	(a)	52.	(b)	53.	(c)	54.	(c)	55.	(a)	56.	(a)
57.	(c)	58.	(b)	59.	(d)	60.	(b)	61.	(d)	62.	(c)	63.	(d)	64.	(a)
65.	(a)	66.	(b)	67.	(c)	68.	(b)	69.	(c)	70.	(d)	71.	(b)	72.	(d)
73.	(b)	74.	(d)	75.	(d)	76.	(c)	77.	(b)	78.	(d)	79.	(b)	80.	(d)
81.	(b)	82.	(c)	83.	(c)	84.	(b)	85.	(a)	86.	(a)	87.	(b)	88.	(c)
89.	(d)	90.	(d)	91.	(d)	92.	(c)	93.	(b)	94.	(b)				

SHORT QUESTIONS

Q.1 What is classification?
Q.2 What are the main objectives of classification?
Q.3 Explain the simple classification.
Q.4 Write four important bases of classification of data.
Q.5 Write down the different kinds of classification.
Q.6 Mention the different characteristics of classification.
Q.7 Describe the modes of classification.
Q.8 Differentiate between inclusive classification and exclusive classification.
Q.9 Distinguish between one-way and two-way classification.
Q.10 How do you find the class interval in the construction of a frequency distribution?
Q.11 Write down the formulas for the number of classes and class interval for a frequency distribution.
Q.12 Define the terms class limits and class boundaries.
Q.13 What are the open-end classes?
Q.14 What is meant by a frequency distribution?
Q.15 Write down the main steps involved in the construction of a frequency distribution from individual obs

Q.16 Explain the construction of a frequency distribution.

Q.17 Write down a sequence of steps that can be followed to assist in preparing a frequency distribution that provides a good representation of the raw scores.

Q.18 Differentiate between grouped data and ungrouped data.

Q.19 Define relative frequency distribution.

Q.20 What is meant by cumulative frequency distribution?

Q.21 Write a short note on cumulative frequency distribution.

Q.22 Explain the percentage frequency distribution.

Q.23 Define a bivariate frequency distribution.

Q.24 What is the difference between univariate frequency distributions and bivariate frequency distributions?

Q.25 Write a short note on cross table or contingency table.

Q.26 Define the term tabulation.

Q.27 What is meant by tabulation of statistical data?

Q.28 Write down the general rules of tabulation.

Q.29 Write down the important points to prepare a good table.

Q.30 Describe the main parts of the table.

Q.31 Differentiate between box-head and stub.

Q.32 Define the terms pre-factory note and footnote.

Q.33 Name the different types of diagrams used in representing data.

Q.34 What are the advantages of diagrams?

Q.35 Write down the name of diagrams which are one-dimensional.

Q.36 Write down the formula of the angle of a sub-sector used in pie chart.

Q.37 Define component bar chart.

Q.38 Write down the main steps in the construction of pie diagram or sector diagram.

Q.39 Explain the multiple bar diagrams.

Q.40 Differentiate between a component bar chart and a pie chart.

Q.41 Define the skewness.

Q.42 What is histogram?

Q.43 Define frequency polygon.

Q.44 Write short note on cumulative frequency polygon.

Q.45 Write down the important graphs of frequency distributions.

Q.46 What are the merits of graphs?

Q.47 Distinguish between histogram and historigram.

Q.48 Write any six rules for drawing graphs.

Q.49 Define the term ogive.

Q.50 Explain the symmetry.

Q.51 Name the different types of diagrams which are two-dimensional

Q.52 Distinguish between diagrammatic and graphic representation.

Q.53 Differentiate between diagrams and graphs. Write names of graphs of frequency distributions.

EXERCISES

Q.1 Array the following raw data from a sample of n = 10 midterm examination scores in statistics:

63, 45, 95, 83, 90, 78, 60, 88, 93, 54

Q.2 The following values relate to the number of members in various families. Make a frequency distribution, taking one as the size of interval.

8	9	9	7	6	5	4	2	3	4	6	5	6	11	1	6	9	8	7	6
2	10	9	7	6	8	6	6	5	7	9	4	3	2	8	5	7	5	6	4
5	6	3	5	7	7	10	11	2	1	4	5	3	5	6	7	6	10	6	5

Q.3 The following are the number of flowers on different branches of a tree. Make a discrete frequency distribution, taking a class interval of size one.

3	2	2	1	2	1	2	10	4	0	1	2	3	9	4	8	10	1	2	9	6	7	5	2	4
6	3	10	9	9	5	1	0	5	1	2	4	4	6	7	4	6	8	7	5	3	3	1	6	4

Q.4 The following data shows the number of heads in an experiment of 50 sets of tossing a biased coin 5 times. Make a discrete frequency distribution.

3	2	1	5	2	1	2	3	1	2	4	1	2	3	1	2	0	1	4	2	3	2	1	2	5
5	2	4	2	1	4	4	1	2	2	4	1	2	3	2	1	1	3	0	2	2	2	3	4	5

Q.5 Read the following passage and observe the word length of each word. Make a frequency distribution of word length.

"Statistics is very important to the banks. The banks do business with the help of deposits in the banks. The banks do not keep all the deposits with them. They do business with it and lend the money to the borrowers. The banks work on the assumption that all depositors do not withdraw their deposits on the same day. The bankers use statistical approach based on probability to estimate the number of depositors and their claims for a certain day."

Q.6 Tabulate the following marks into a frequency distribution taking 10 as class interval by inclusive method and 45 as the lowest limit. Make the frequency polygon for the frequency distribution so obtained.

109	74	103	95	90	118	52	88	101	96	72	56	64	110	97	59	62	96
82	62	85	105	116	91	83	99	52	76	84	89	77	104	101	107	62	46

Q.7 The following is a record of weights of 70 students (in lbs.). Tabulate the data in the form of frequency distribution taking the classes like 60 – 64, 65 – 69,.

72	96	69	78	76	112	107	93	73	61
75	91	106	84	84	103	109	96	88	80
86	90	105	77	90	101	91	102	92	91
82	99	63	95	118	77	72	114	101	113
86	83	76	111	107	92	89	87	106	100
93	98	85	115	108	73	94	62	107	106
92	88	88	104	[illegible]	67	98	74	97	109

Q.8 The following data give the index numbers of 50 commodities in a certain year. Tabulate the data into a frequency distribution taking a class interval of size 5.

120	138	113	119	111	118	123	151	117	134
141	115	119	124	112	145	114	114	106	113
110	111	106	153	100	101	108	133	100	109
110	143	109	138	113	136	100	104	116	144
128	105	106	147	127	129	140	120	129	108

Q.9 Form a frequency distribution of the following taking 6 as class interval by exclusive method.

79.6	72.1	64.2	82.9	68.9	74.2	81.8	90.6
55.9	71.4	71.6	63.5	88.1	77.8	74.2	80.7
65.7	75.2	95.0	59.4	58.3	69.4	83.2	82.7
73.8	67.6	81.9	73.5	77.6	48.6	83.5	70.8

Q.10 Make a frequency distribution taking the classes as 1.19 – 1.23, 1.24 – 1.28, etc., from the following data:

1.35	1.55	1.73	1.76	1.48	1.32	1.32	1.76
1.47	1.46	1.68	1.60	1.63	1.36	1.45	1.47
1.26	1.59	1.64	1.46	1.50	1.19	1.48	1.24
1.61	1.38	1.41	1.50	1.40	1.45	1.56	1.51
1.65	1.45	1.67	1.54	1.64	1.35	1.49	1.45

Also make the class boundaries.

Q.11.(i) The grades in Mathematics of 50 students are as under:

68	76	71	60	82	96	83	76	78	73
93	59	75	71	65	78	81	78	73	95
74	71	88	82	62	75	76	63	88	61
94	53	90	73	65	72	97	74	68	75
66	75	85	88	60	69	85	57	67	77

Form a frequency distribution taking intervals of 5 grades like 50 – 54, 55 – 59, etc.

(ii) Draw a histogram of the grouped data formed in part (i) on graph paper.

Q.12 The following table shows the diameters in millimetres of a sample of 60 ball bearings manufactured by a company. Convert to a frequency table with 6 classes, beginning with 7.24 by using inclusive classification.

7.40	7.36	7.34	7.35	7.35	7.34	7.27	7.31	7.44	7.26
7.41	7.42	7.32	7.36	7.40	7.35	7.35	7.39	7.39	7.25
7.34	7.24	7.35	7.32	7.32	7.34	7.33	7.38	7.38	7.28
7.40	7.42	7.35	7.27	7.36	7.36	7.33	7.39	7.30	7.36
7.37	7.30	7.31	7.29	7.41	7.35	7.46	7.43	7.32	7.29
7.35	7.37	7.33	7.36	7.45	7.30	7.28	7.32	7.38	7.37

Q.13 The following table gives the weights in pounds of 40 students at a college. Construct a frequency distribution of the weights using appropriate class interval.

128, 150, 156, 145, 147, 165, 142, 140, 154, 135, 158, 118, 145, 146,
163, 161, 157, 180, 135, 149, 138, 140, 125, 126, 153, 144, 168, 135,
132, 144, 147, 150, 152, 142, 164, 148, 173, 138, 136, 146.

Q.14 The following data represent reported sales (in millions of rupees) for 26 companies in the shoe industry:

32, 36, 54, 38, 17, 41, 22, 33, 22, 32, 31, 21, 18,
46, 36, 11, 31, 29, 12, 23, 51, 12, 13, 37, 33, 27.

Construct a frequency distribution using classes with a width of 10 i.e; 10 – 20, 20 – 30, etc.

Q.15 The class marks for the ages of sales clerks employed in a department store are 18.5, 28.5, 38.5, 48.5, 58.5 and 68.5. Find the class boundaries of this distribution.

Q.16 The daily expenditure (rupees) in an office are grouped into a table having the classes 100 – 149, 150 – 199, 200 – 249, 250 – 299 and 300 – 349.

Find (i) the class boundaries (ii) the class marks (iii) the class interval

Q.17 Write down the class boundaries, class marks and class intervals for each of the following classes.

(a) 4 – 9, 10 – 15 (b) 2.1 – 2.4, 2.5 – 2.8

(c) – 3 – (+ 3), 4 – 10 (d) 8, 12, 16

Q.18 Given the frequencies of 5 classes 2, 6, 8, 4, 5. Find relative frequencies and cumulative frequencies.

Q.19 The following are the shipping weights in pounds of 60 supposedly identical office desks.

Shipping weight	98-98.9	99-99.9	100-100.9	101-101.9	102-102.9
Number of desks	7	23	14	13	3

Find: (i) the class boundaries (ii) the class marks (iii) the relative frequency distribution (iv) the cumulative percentage frequency distribution

Q.20 Consider the following data:

Hourly wages (Rs.)	30 – 49	50 – 69	70 – 89	90 – 109	110 – 129	130 – 149
Number of persons	4	20	23	35	10	8

Determine: (i) Relative frequency distribution
(ii) Percentage frequency distribution
(iii) More than cumulative frequency distribution
(iv) Less than cumulative frequency distribution
(v) Cumulative percentage frequency distribution.

Q.21 The masses measured to the nearest kg. of 50 boys are noted and the distribution formed.

Masses (kg.)	60 – 64	65 – 69	70 – 74	75 – 79	80 – 84	85 – 89
Number of boys	2	6	12	14	10	6

Construct relative frequency, percentage frequency, cumulative relative frequency and cumulative percentage frequency distributions.

Q.22 Given the following frequency distribution, fill in the missing frequencies:

Hourly wages (Rs.)	No. of workers (f)	Cumulative frequency (c.f)
50 – 70	10	10
70 – 90	20	30
90 – 110	24	–
110 – 130	–	–
130 – 150	14	–
150 – 170	8	104

Q.23 The daily wages in rupees are given in the following table. Find the missing values.

Daily wages (Rs.)	No. of workers (f)	Cumulative frequency (c.f)
500 – 519	–	4
520 – 539	20	–
540 – 559	23	47
560 – 579	–	82
580 – 599	10	–
600 – 619	–	–
Total	100	

Q.24 The daily wages in rupees are given in the following table. Find the missing values.

Daily Wages (Rs.)	No. of Workers	Relative Frequency	Cumulative Frequency
400 – 419	—	—	—
420 – 439	20	—	24
440 – 459	—	0.23	—
460 – 479	—	0.35	82
480 – 499	10	—	—
500 – 519	—	—	—
Total	100	—	—

Q.25 Following are the monthly salaries of a sample of 50 workers in an office. Determine the missing values.

Monthly Salary (in thousand Rs.)	No. of workers (f)	Relative Frequency	Cumulative Frequency
16 and under 18	—	0.12	—
18 and under 20	5	—	11
20 and under 22	—	0.20	21
22 and under 24	20	—	—
24 and under 26	—	0.10	46
26 and under 28	4	—	—

Q.26 The masses measured in kilogrammes of 100 students are noted and the distribution formed. Construct a frequency distribution of data.

Masses (kg.)	X < 59.5	X < 62.5	X < 65.5	X < 68.5	X < 71.5	X < 74.5
No. of Students	0	5	23	65	92	100

Q.27 Consider the following data:

Hourly wages (Rs.)	X > 50	X > 60	X > 70	X > 80	X > 90	X > 100
"More than" Cumulative frequency	100	96	76	53	18	8

Make a frequency distribution. Draw a histogram.

Q.28 Make a bivariate frequency distribution for the data given below, taking the classes of heights such as 60 – 62, 63 – 65 etc. and the classes of weights as 100 – 104, 105 – 109 etc.

Heights (inches)	Weights (pounds)	Heights (inches)	Weights (pounds)	Heights (inches)	Weights (pounds)
60	100	61	109	63	108
62	105	60	108	67	108
61	104	63	107	71	116
70	115	64	112	70	110
64	110	67	115	68	114
60	102	68	117	68	116
65	110	69	117	71	119
65	108	64	111	64	107
73	119	66	113	63	108
71	118	62	104	68	105
73	119	69	107	64	115
67	111	67	114	64	108
62	105	67	117	62	105
64	107				

Q.29 The following table shows the heights (cm.) and weights (kg.) of 60 students. Make a bivariate frequency distribution, taking a class interval of 5 cm. for heights and 3 kg. for weights.

Heights (cm.)	170	162	155	158	155	170	168	150	160	158	160	178
Weights (kg.)	52	54	52	54	52	58	58	50	54	54	56	58
Heights (cm.)	168	172	158	168	178	150	178	165	168	152	165	162
Weights (kg.)	59	54	54	56	58	54	59	56	58	60	56	54
Heights (cm.)	162	160	170	160	152	165	162	169	168	175	170	160
Weights (kg.)	54	54	57	56	52	58	55	54	57	55	58	55
Heights (cm.)	170	160	160	172	150	152	168	152	172	155	156	155
Weights (kg.)	58	58	54	58	51	52	54	54	58	52	60	58
Heights (cm.)	175	168	170	178	175	174	165	159	172	168	166	176
Weights (kg.)	60	56	62	63	64	66	56	60	65	67	62	64

Q.30 The areas of the various continents/country of the world in millions of square kilometers are given as:

Continent / country	Area (million of square kilometers)
Africa	30.3
Asia	26.9
Europe	4.9
North America	24.3
Oceania	8.5
South America	17.9
U.S.S.R.	20.5

Graph the data by (i) A simple bar chart. (ii) A pie chart.

Q.31 The following table gives the details of monthly budgets of two families. Represent these figures by a suitable diagram.

Items	Family A	Family B
Food	Rs.6000	Rs.8000
Clothing	Rs.1000	Rs.1000
House rent	Rs.4000	Rs.5000
Fuel and lighting	Rs.1000	Rs.1000
Miscellaneous	Rs.3000	Rs.5000
Total	Rs. 15000	Rs. 20000

Q.32 Draw a component bar chart to represent the following data:

	Expenditure in thousands of rupees		
Year	Agriculture	Industries	Miscellaneous
1999	600	245	222
2000	675	263	230
2001	850	320	255
2002	825	375	300
2003	900	425	350

Q.33 The following table gives the details of monthly expenditure (in thousand rupees) of three families. Draw a sub divided rectangular diagram:

	Family		
Items of expenditure	A	B	C
Food	16	21	24
Clothing	5	9	8
House rent	10	12	20
Education	8	6	12
Recreation	3	3	4
Conventional needs	2	3	4
Miscellaneous	6	6	8

Q.34 Represent the following data by pie diagram:

Items of expenditure	Amount
Food	12000
Clothing	3000
House rent	6000
Education	4000
Fuel and light	2000
Miscellaneous	3000

Q.35 Given below is the area under wheat in Rawalpindi Division by districts in 1989 – 90:

Districts	Attock	Chakwal	Jhelum	Rawalpindi
Area (Hectares)	162	135	56	135

Represent the data by a pie chart.

Source: Directorate of Agriculture Crop Reporting Service, Punjab, Lahore.

Q.36 The following table gives the details of monthly expenditure (in thousand rupees) of three families. Construct a pie diagram to compare the budgets of three families A, B and C.

Item of Expenditure	Family A	Family B	Family C
Food	12	15	20
Clothing	3	4	5
House Rent	6	8	10
Education	4	5	6
Fuel	2	3	5
Miscellaneous	3	5	4

Q.37 The following are daily numbers of automobiles rented by an automobile rental company in 90 business days.

Automobile rentals	20-24	25-29	30-34	35-39	40-44	45-49	50-54
Number of days	3	10	21	28	14	9	5

Draw histogram and frequency polygon of this distribution.

Q.38 Draw histogram and frequency polygon on the graph paper for the following frequency distribution.

Hourly wages (Rs.)	50-54	55-59	60-64	65-69	70-74	75-79	80-84
Number of workers	4	8	12	20	16	10	5

Q.39 The following table gives the weights in pounds of 40 students at a college. Draw a cumulative frequency polygon.

Weights (pounds)	118 – 126	127 – 135	136 – 144	145 – 153	154 – 162	163 – 171	172 – 180
No. of students	3	5	9	12	5	4	2

Q.40 The daily wages of 150 workers are given below in the form of grouped data. Draw frequency polygon, frequency curve and ogive polygon.

Daily wages (Rs.)	No. of workers (f)	Daily wages (Rs.)	No. of workers (f)
400 – 450	15	600 – 650	20
450 – 500	25	650 – 700	10
500 – 550	30	700 – 750	5
550 – 600	45	Total	150

Chapter 3

MEASURES OF CENTRAL TENDENCY

3.1. INTRODUCTION

In our daily life, it is quite common to hear the following statements.

(i) The price of apples in the market is Rs.35 per kg.

(ii) Mr. A studies for 8 hours daily.

(iii) Death rate in Pakistan is 15 per thousand.

(iv) Consumption of meat in a family is 5 kg. per month.

(v) The rain fall in a country is 25 cms. per year.

(vi) Mr. B plays cricket daily for 2 hours.

If we concentrate on the above statements, we find that none of them is the exact statement. The prices of apples are generally different on different shops. On some particular day the prices may change from morning to evening. Statement (ii) is about the study hours of Mr. A. May be that Mr. A has never studied for exactly 8 hours. His study hours are approximately 8 hours. These statements are not exactly true but still they are very important. We can say that these are approximate statements about different situations. In statistics we say that these are the average statements. In our daily conversation, we make many statements which have some meaning only on average basis. When we talk about some numerical data or some situation is described in a numerical manner, the word average plays very important role. In different fields of life like agriculture, education, industry, prices etc., the idea of average is very important. There are very large number of experts at national and international level who are involved in the study of averages of various kinds. Some experts are interested to find the average production of wheat per acre in different countries, some others may be interested in the average prices of wheat in those countries. Thus the word average is deeply linked with the situations which need some numerical study or analysis. The average is also called measure of central tendency.

3.2. CALCULATION OF AVERAGE

Average is a single value which is calculated to represent the whole data. It may be calculated for a sample or a population data. The average is a value which expresses the central idea of the observations. It is a single value used to represent the data. It is a value somewhere in the centre, where most of the items of the series cluster. Such values are also called measures of central tendency. Different averages are used in different situations to represent the data. A choice of the proper average is the job of an expert who is calculating the average.

3.3. QUALITIES OF A GOOD AVERAGE

Following qualities are associated with a good average. We try to use that average which possesses all or most of these qualities.

(i) It should be rigidly defined. ~~It means that there should be no confusion about~~ its calculation.

(ii) It should be based on all the observations of the data.

(iii) It should be capable of further algebraic treatment.

(iv) It should be unaffected by extreme observations.

(v) It should be easy to calculate and simple to follow.

(vi) It should be least affected by fluctuations of sampling.

3.4. TYPES OF AVERAGES

The following averages are usually used:

(1) Arithmetic Mean. (2) Median, Quartiles and other partition values.

(3) Mode. (4) Geometric Mean.

(5) Harmonic Mean

3.5. ARITHMETIC MEAN

The arithmetic mean of a set of n observations is defined as the total of all the observations divided by the number of observations. The arithmetic mean of sample data is denoted by $\bar{X}$ and the symbol μ (mue) is used for mean of population data.

3.5.1. INDIVIDUAL OBSERVATIONS (UNGROUPED DATA)

Let the variable X takes the values, $X_1, X_2, X_3, \ldots, X_n$, then the arithmetic mean, denoted by $\bar{X}$ (read "X bar") is defined by

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{\Sigma X_i}{n}$$

where, $\bar{X}$ = arithmetic mean, ΣX_i = sum of all the items of the variable X,

n = number of observations.

Example 3.1.

Find the arithmetic mean of the set of numbers 84, 91, 72, 68, 87 and 78.

Solution:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = \frac{480}{6} = 80$$

Example 3.2.

Suppose that we are interested in examining how much a computer system is being used by various departments. Compute the arithmetic mean for the number of uses of the computer system from the following data.

Department number	1	2	3	4	5	6	7
No. of uses per month	20	65	120	280	350	325	310

Solution:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{20 + 65 + 120 + [illegible] + 350 + 325 + 310}{7} = \frac{1470}{7} = 210$$

Example 3.3.

The reciprocals of the values of the variable X are 0.0500, 0.0400, 0.0200, 0.0285, 0.0143. Find arithmetic mean of X.

Solution: The necessary calculations are given below:

1/X	0.0500	0.0400	0.0200	0.0285	0.0143
X	20	25	50	35	70

Here, $n = 5$, $\Sigma X = 200$, $\bar{X} = \frac{\Sigma X}{n} = \frac{200}{5} = 40$

Example 3.4.

The logarithms of 10 values of X are 1.8062, 1.2304, 1.6532, 1.5798, 1.4314, 0.7782, 1.6812, 1.0414, 1.7559, 1.5315. Calculate arithmetic mean of X values.

Solution: The necessary calculations are given below:

log X	1.8062	1.2304	1.6532	1.5798	1.4314	0.7782	1.6812	1.0414	1.7559	1.5315
X	64	17	45	38	27	6	48	11	57	34

Here, $n = 10$, $\Sigma X = 347$, $\bar{X} = \frac{\Sigma X}{n} = \frac{347}{10} = 34.7$

Example 3.5.

The mean salary paid to 1000 employees of a factory was found to be Rs.180.40. Later on it was discovered that the wages of two employees were wrongly taken as 297 and 165 instead of 197 and 185. Find the correct mean.

Solution:

$$\bar{X} = \frac{\Sigma X}{n} \quad \text{or} \quad \Sigma X = n\bar{X}$$

$$\begin{aligned}
\text{Incorrect } \bar{X} &= 180.40, \text{ Incorrect } \Sigma X = 1000(180.40) = 180400 \\
\text{Correct } \Sigma X &= \text{Incorrect } \Sigma X - \text{Wrong items} + \text{Correct items} \\
&= 180400 - 297 - 165 + 197 + 185 = 180320 \\
\text{Correct } \bar{X} &= \frac{\text{Correct } \Sigma X}{n} = \frac{180320}{1000} = \text{Rs.}180.32
\end{aligned}$$

Example 3.6.

(a) The mean of 15 values is 10. If one more value is included, the mean becomes 12. Find the value which is included.

(b) The mean of n values is 8. If a new value 28 is included, the mean becomes 9. Find the value of n.

Solution:

(a) $\bar{X} = \frac{\Sigma X}{n} = \frac{\Sigma X}{15} = 10 \quad \text{or} \quad \Sigma X = 10(15) = 150$

Let A denote the value which is included, then mean of 16 values is

$\frac{\Sigma X + A}{n+1} = \frac{\Sigma X + A}{15+1} = 12 \quad \text{or } \Sigma X + A = 12(16) = 192 \text{ or}$

$A = 192 - \Sigma X = 192 - 150 = 42$

Hence, the included value is 42.

(b) $\bar{X} = \frac{\Sigma X}{n} = 8 \quad \text{or} \quad \Sigma X = 8n$

When 28 is included, the total of $(n + 1)$ values becomes $\Sigma X + 28$.

Thus, mean of $n + 1$ values $= \frac{\Sigma X + 28}{n+1} = 9$ or $\Sigma X + 28 = 9(n + 1) = 9n + 9$

or $\quad \Sigma X = 9n + 9 - 28 = 9n - 19$

$\quad 8n = 9n - 19 \qquad (\text{since } \Sigma X = 8n)$

or $\quad 9n - 8n = 19 \quad \text{or} \quad n = 19$

3.5.2. FREQUENCY DISTRIBUTION (DISCRETE DATA)

If $f_1, f_2, f_3, \ldots, f_n$ be the frequencies of $X_1, X_2, X_3, \ldots, X_n$ respectively, then the arithmetic mean is defined by

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \ldots + f_nX_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\Sigma f_iX_i}{\Sigma f_i} \text{ or } \frac{\Sigma f_iX_i}{n}$$

where Σf_iX_i = Sum of the products of the frequencies and the variable X.

$$n = f_1 + f_2 + f_3 + \ldots + f_n = \Sigma f_i = \text{Total frequency.}$$

Example 3.7.

Given below is the frequency distribution of number of apples on 100 apple trees. Find the average number of apples per tree.

Number of apples (X)	500	550	590	620	700	740
Number of trees (f)	15	25	30	20	7	3

Solution:

The necessary calculations are given below:

X	500	550	590	620	700	740	Total
f	15	25	30	20	7	3	100
fX	7500	13750	17700	12400	4900	2220	58470

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{58470}{100} = \text{[illegible]} 7$$

3.5.3. GROUPED DATA

Suppose the observations are put into n different classes and the mid points of the classes are calculated which represent all the observations of the respective class. If the mid points of the classes are X_1, X_2, X_3, ..., X_n with corresponding frequencies f_1, f_2, f_3, ..., f_n, then the arithmetic mean $\bar{X}$ is defined as

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \ldots + f_nX_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\Sigma f_iX_i}{\Sigma f_i}$$

Example 3.8.

The following frequency distribution shows the hourly income of 100 households in a locality.

Income (Rs.)	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64	65 – 69
Frequency	13	15	28	17	12	10	5

Calculate the arithmetic mean and show that sum of the deviations of values from their mean is zero.

Solution:

The necessary calculations are given below:

Income	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64	65 – 69	Total
X	37	42	47	52	57	62	67	--
f	13	15	28	17	12	10	5	100
fX	481	630	1316	884	684	620	335	4950
$X-\bar{X}$	– 12.5	– 7.5	– 2.5	2.5	7.5	12.5	17.5	--
$f(X-\bar{X})$	– 162.5	– 112.5	– 70.0	42.5	90.0	125.0	87.5	0

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{4950}{100} = 49.5. \text{ Hence } \Sigma f(X-\bar{X}) = 0$$

3.5.4. WEIGHTED ARITHMETIC MEAN

Sometimes different observations do not have the same importance. Some observations, for some reason, have greater importance. The relative importance called the weight (W) of the observations is thus determined. If the observations X_1, X_2, X_3, ..., X_n have the respective weights W_1, W_2, W_3, ..., W_n, the weighted arithmetic mean denoted by $\bar{X}_w$ is defined as

$$\bar{X}_w = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots + W_nX_n}{W_1 + W_2 + W_3 + \ldots + W_n} = \frac{\Sigma WX}{\Sigma W}$$

Example 3.9.

A student's final marks in Mathematics, Physics, English and Statistics are respectively 82, 86, 90 and 70. If the respective credits received for these courses are 3, 5, 3 and 1, determine an approximate average marks.

Solution:

The necessary calculations are given below:

X	82	86	90	70	Total
W	3	5	3	1	12
WX	246	430	270	70	1016

$$\text{Weighted Arithmetic Mean} = \bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{1016}{12} = 84.67 \text{ or } 85$$

3.5.5. COMBINED ARITHMETIC MEAN

If $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k$ be the arithmetic mean of k distributions with respective frequencies $n_1, n_2, n_3, \ldots, n_k$, the combined arithmetic mean $\bar{X}_c$ is defined by

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3 + \ldots + n_k\bar{X}_k}{n_1 + n_2 + n_3 + \ldots + n_k} = \frac{\Sigma n_i\bar{X}_i}{\Sigma n_i}$$

Example 3.10.

Find out combined mean from the following data:

	Series X_1	Series X_2
Arithmetic mean	12	20
No. of items	80	60

Solution:

Here, $n_1 = 80, n_2 = 60, \bar{X}_1 = 12$ and $\bar{X}_2 = 20$. Therefore

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{80(12) + 60(20)}{80 + 60} = \frac{2160}{140} = 15.43$$

Example 3.11

In three separate weeks a discount store sold 10, 25 and 15 microwave ovens at average prices of Rs. 8000, Rs. 6000, and Rs. 6500 respectively. What is the average price of the ovens sold?

Solution:

Here $n_1 = 10, n_2 = 25, n_3 = 15, \bar{X}_1 = 8000, \bar{X}_2 = 6000$ and $\bar{X}_3 = 6500$. Therefore

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1 + n_2 + n_3} = \frac{10(8000) + 25(6000) + 15(6500)}{10 + 25 + 15}$$

$$= \frac{327500}{50} = 6550$$

Example 3.12.

(i) Average income of a group of 10 workers is Rs.49 per day. Another group with average income of Rs.60 per day has joined the first group. Both the groups have agreed to distribute their total income equally which comes to Rs.55. Find the number of workers in the second group.

(ii) The average wage of 4 men is Rs.17 per week. What is the average wage of further 6 men if the average wage of all 10 men is Rs.20?

(iii) The mean marks of 80 students in a certain class was 65. The mean marks of male students is 70 and that of female students is 50. Find the number of boys and the number of girls in the class.

Solution:

(i) Let n_2 be the number of workers in the second group

We have, $\bar{X}_c = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$

Here, $n_1 = 10, \ \bar{X}_1 = 49, \ \bar{X}_2 = 60, \bar{X}_c = 55, n_2 = ?$

Substituting the values, we get

$$55 = \frac{10(49) + 60\,n_2}{10 + n_2} \text{ or } 490 + 60\,n_2 = 55(10 + n_2) \text{ or}$$

$$490 + 60\,n_2 = 550 + 55\,n_2 \text{ or } 60\,n_2 - 55\,n_2 = 550 - 490 \text{ or}$$

$$5\,n_2 = 60 \text{ or } n_2 = 12 \text{ workers}$$

(ii) We have, $\bar{X}_c = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$

Here, $\bar{X}_c = 20, n_1 = 4, \ \bar{X}_1 = 17, n_2 = 6, \ \bar{X}_2 = ?$

Substituting the values in the above formula, we get

$$20 = \frac{4(17) + 6\bar{X}_2}{4 + 6} = \frac{68 + 6\bar{X}_2}{10} \text{ or } 68 + 6\bar{X}_2 = 20(10) = 200 \text{ or}$$

$$6\bar{X}_2 = 200 - 68 = 132 \text{ or } \bar{X}_2 = 22$$

Hence the average wage of 6 men is Rs.22.

(iii) We have, $\bar{X}_c = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$ and $n_1 + n_2 = n$

Here, $\bar{X}_c = 65, \bar{X}_1 = 70, \ \bar{X}_2 = 50, \ n = 80, \ n_1 = ?, \ n_2 = ?$

Substituting the values, we get

$$65 = \frac{70\,n_1 + 50\,n_2}{80} \text{ or } 70\,n_1 + 50\,n_2 = 80(65) = 5200 \quad \cdots\cdots \quad (1)$$

$$n_1 + n_2 = 80 \quad \cdots\cdots \quad (2)$$

Solving equations (1) and (2), we get $n_2 = 20$

Substituting $n_2 = 20$ in equation (2), we get $n_1 + 20 = 80$ or $n_1 = 80 - 20 = 60$

3.5.6. CHANGE OF ORIGIN

When a certain constant, say A is added to all the observations, we get a new set of observations. Similarly, the observations can be decreased when a constant is subtracted from all the observations. The addition or subtraction of a constant is called change of origin. This constant may be called arbitrary origin or working mean.

Let $X_i - A = D_i$, where A is a constant and D_i is the new variable, then

$$\Sigma(X_i - A) = \Sigma D_i \quad \text{or} \quad \Sigma X_i - \Sigma A = \Sigma D_i$$

$$(\Sigma A = A + A + A + \ldots \text{ n times } = nA)$$

$$\Sigma X_i - nA = \Sigma D_i$$

Dividing both sides of the equation by n, we get

$$\frac{\Sigma X_i - nA}{n} = \frac{\Sigma D_i}{n} \quad \text{or} \quad \frac{\Sigma X_i}{n} - \frac{nA}{n} = \frac{\Sigma D_i}{n} \quad \text{or}$$

$$\bar{X} - A = \bar{D} \quad \text{or} \quad \bar{X} = A + \bar{D} = A + \frac{\Sigma D_i}{n}$$

For any frequency distribution, $\bar{X} = A + \dfrac{\Sigma f_i D_i}{\Sigma f_i}$

This relation can also be used for the calculation of $\bar{X}$. Some people call it short method. If the size of the observations is very large, the computational work can be reduced by using the idea of change of origin. But we shall not suggest our students to use this method in general. Similarly if $X_i + A = D_i$, then $\bar{X} + A = \bar{D}$ or $\bar{X} = \bar{D} - A$. This equation is only of academic interest and $\bar{X}$ is never calculated through this relation.

Example 3.13.

The wages of 5 workers in rupees are 1950, 2000, 2050, 2060, 2080. Calculate the arithmetic mean by using the idea of change of origin (short cut method) .

Solution:

Let us take 2000 as the constant, then

X	1950	2000	2050	2060	2080	Total
D = X – 2000	– 50	0	50	60	80	140

Thus, $\bar{X} = A + \dfrac{\Sigma D}{n} = 2000 + \dfrac{140}{5} = 2000 + 28 = \text{Rs.}2028$

Example 3.14.

The deviations of the data about X = 180 are 4, 11, – 8, – 12, 7, 9, 16, 9, 13, 15. Calculate the arithmetic mean.

Solution:

$D = X - 180$	4	11	-8	-12	7	9	16	9	13	15

Here, $\Sigma D = 64$, $n = 10$, $A = 180$, $\bar{X} = A + \dfrac{\Sigma D}{n} = 180 + \dfrac{64}{10} = 186.4$

Example 3.15.

Given below is the grouped data of wages of 500 workers of a factory. Calculate the arithmetic mean by short method.

Groups	1900 – 1950	1950 – 2000	2000 – 2050	2050 – 2100	2100 – 2150
No. of workers	60	180	185	65	10

Solution:

Let us take A = 2025 as the arbitrary origin which corresponds to the maximum frequency. The necessary calculations are given below:

Groups	1900 – 1950	1950 – 2000	2000 – 2050	2050 – 2100	2100–2150	Total
Mid Point (X)	1925	1975	2025	2075	2125	--
No. of workers (f)	60	180	185	65	10	$\Sigma f = 500$
D = X – 2025	– 100	– 50	0	50	100	--
fD	– 6000	– 9000	0	3250	1000	$\Sigma fD = -10750$

Thus, $\bar{X} = A + \dfrac{\Sigma fD}{\Sigma f} = 2025 + \dfrac{-10750}{500} = 2025 - 21.5 = 2003.5$

3.5.7. CHANGE OF SCALE

When all the observations are divided by a certain constant or multiplied by a constant, the process is called change of scale. Let us take a variable u_i called a coded variable, where $u_i = \dfrac{X_i}{h}$, h being any constant. We have $X_i = hu_i$

Taking summation of both sides, we have

$$\Sigma X_i = \Sigma hu_i = h\Sigma u_i$$

Dividing both sides by n, we get

$$\frac{\Sigma X_i}{n} = h\,\frac{\Sigma u_i}{n} \quad \text{or} \quad \bar{X} = h\bar{u}$$

This formula can also be used for the calculation of $\bar{X}$. Similarly if $u_i = hX_i$, then $\bar{u} = h\bar{X}$ or $\bar{X} = \bar{u}/h$

3.5.8. CHANGE OF ORIGIN AND SCALE

When both the operations of addition (or subtraction) and multiplication (or division) with a constant are applied on the observations, we get a new variable u_i,

called the coded variable. Thus the values of $u_i = \frac{X_i - A}{h}$, $u_i = h(X_i - A)$, $u_i = \frac{X_i + A}{h}$ and $u_i = h(X_i + A)$ are the result of the operations of change of origin and scale.

If $u_i = \frac{X_i - A}{h}$, then $X_i = A + hu_i$

Applying summation on both sides

$$\Sigma X_i = \Sigma(A + hu_i) = nA + h\Sigma u_i$$

Dividing both sides by n, we get

$$\frac{\Sigma X_i}{n} = \frac{nA + h\Sigma u_i}{n} = \frac{nA}{n} + h\frac{\Sigma u_i}{n} \quad \text{Thus} \quad \bar{X} = A + h\bar{u}$$

This formula can be used for the calculation of $\bar{X}$. For frequency distribution the relation will be

$$\bar{X} = A + h\frac{\Sigma fu}{\Sigma f} = A + h\bar{u}$$

It may be noted that if there is some relation between two variables say X_i and u_i, the same relation exists between their means. If

$$u_i = \frac{X_i - A}{h}, \text{ then } \bar{u} = \frac{\bar{X} - A}{h}, \text{ if } u_i = h(X_i - A), \text{ then } \bar{u} = h(\bar{X} - A)$$

Example 3.16.

The weights of 150 students are given below in the form of grouped data. Calculate the mean by use of "change of origin and scale".

Weights (lbs.)	95–100	100–105	105–110	110–115	115–120	120–125	125–130
Frequency	7	17	37	35	28	15	11

Solution:

The necessary calculations are given below:

Weights (lbs.)	X	f	A = 107.5, h = 5 $u = \frac{X - 107.5}{5}$	fu
95 – 100	97.5	7	−2	−14
100 – 105	102.5	17	−1	−17
105 – 110	107.5	37	0	0
110 – 115	112.5	35	1	35
115 – 120	117.5	28	2	56
120 – 125	122.5	15	3	45
125 – 130	127.5	11	4	44
Total		150		149

$$\bar{X} = A + \frac{\Sigma fu}{\Sigma f} \times h = 107.5 + \frac{149}{150} \times 5 = 107.5 + 4.97 = 112.47$$

3.5.9. PROPERTIES OF ARITHMETIC MEAN

Some of the mathematical properties of the arithmetic mean are:

(i) The sum of the deviations of the values X_i from their mean $\bar{X}$ is zero i.e.

$$\Sigma(X_i - \bar{X}) = \Sigma X_i - n\bar{X} = n\bar{X} - n\bar{X} = 0 \quad \text{(ungrouped data)}$$

$$\Sigma f_i(X_i - \bar{X}) = \Sigma f_i X_i - \bar{X}\Sigma f_i = \bar{X}\Sigma f_i - \bar{X}\Sigma f_i = 0 \quad \text{(grouped data)}$$

(ii) The sum of the squares of the deviations of the values of a variable is least when the deviations are measured from their mean i.e.

$$\Sigma(X_i - \bar{X})^2 \leq \Sigma(X_i - A)^2 \quad \text{(ungrouped data)}$$

$$\Sigma f_i(X_i - \bar{X})^2 \leq \Sigma f_i(X_i - A)^2 \quad \text{(grouped data)}$$

(iii) If $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k$ be the arithmetic means of k distributions with respective frequencies $n_1, n_2, n_3, \ldots, n_k$, the combined mean $\bar{X}_c$ of the whole distribution is given by

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3 + \cdots + n_k\bar{X}_k}{n_1 + n_2 + n_3 + \cdots + n_k}$$

(iv) If $X_1, X_2, X_3, \ldots, X_n$ be the n observations having arithmetic mean $\bar{X}$ and if $Y = aX \pm b$, then $\bar{Y} = a\bar{X} \pm b$, where a and b are any two numbers and $a \neq 0$. This is called linear transformation of X into Y.

Example 3.17

Let a variable X has the values as 40, 50, 60, 80 and 100. Find the mean of X and verify that $\Sigma(X - \bar{X}) = 0$. Also verify that $\Sigma(X - \bar{X})^2 < \Sigma(X - A)^2$ if $A = 60$. Now multiply each number by 2 and subtract 50 to obtain the observations of a new variable Y. Find the mean of Y and show that $\bar{Y} = 2\bar{X} - 50$.

Solution:

The necessary calculations are given below:

X	$X - \bar{X}$	$(X - \bar{X})^2$	$X - A$	$(X - A)^2$	$Y = 2X - 50$
40	-26	676	-20	400	30
50	-16	256	-10	100	50
60	-6	36	0	0	70
80	$+14$	196	$+20$	400	110
100	$+34$	1156	$+40$	1600	150
$\Sigma X = 330$	$\Sigma(X - \bar{X}) = 0$	$\Sigma(X - \bar{X})^2 = 2320$		$\Sigma(X - A)^2 = 2500$	$\Sigma Y = 410$

$\bar{X} = \frac{\Sigma X}{n} = \frac{330}{5} = 66$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{410}{5} = 82$

$2\bar{X} - 50 = 2(66) - 50 = 82$

Hence $\Sigma(X - \bar{X}) = 0$

$\Sigma(X - \bar{X})^2 < \Sigma(X - A)^2$

and $\bar{Y} = 2\bar{X} - 50$

3.5.10. MERITS AND DEMERITS OF ARITHMETIC MEAN

MERITS

(i) It is rigidly defined.

(ii) It is easy to calculate and simple to follow.

(iii) It is based on all the observations.

(iv) It is determined for almost every kind of data.

(v) It is finite and not infinite.

(vi) It is readily put to algebraic treatment.

(vii) It is least affected by fluctuations of sampling.

DEMERITS

(i) The arithmetic mean is highly affected by extreme values.

(ii) It cannot average the ratios and percentages properly.

(iii) It is not an appropriate average for highly skewed distributions.

(iv) It cannot be computed if any item is missing.

(v) The mean sometimes does not coincide with any of the observed value.

3.6. MEDIAN

Median is the middle most value of the n observations arranged in ascending or descending order of magnitude. Half of the observations are above the median and half are below the median. If n is odd, there is one term in the middle which is called median. If n is even, the average of the two central values is called median. For middle term, we can use the formula

$$\text{Median} = \text{Value of} \left(\frac{n+1}{2}\right) \text{th item}$$

In grouped data the number of observations is usually large, therefore we may use $\frac{n+1}{2}$ or $\frac{n}{2}$ for the middle term. It has become a strong convention that for grouped data, $\frac{n}{2}$ is used to get the middle term.

Example 3.18(a).

The wages of 5 workers in rupees are 1800, 1900, 1700, 2000 and 2200. Find the median.

Solution:

After arranging the observations in ascending order, we get

Rs.1700, 1800, 1900, 2000, 2200

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{5+1}{2}\right)\text{th item}$$

$$= \text{Value of 3rd item} = 1900$$

Example 3.18(b).

The minimum temperature in Murree for the first 10 days of March was $-1, -2, 1, 0, 3, 3, 4, 3, 2, 6$. Find the median.

Solution:

After arranging the observations in ascending order, we get

$$-2, -1, 0, 1, 2, 3, 3, 3, 4, 6$$

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{10+1}{2}\right)\text{th item}$$

$$= \text{Value of (5.5) th item} = \frac{\text{5th item + 6th item}}{2} = \frac{2+3}{2} = 2.5$$

3.6.1. MEDIAN FOR FREQUENCY DISTRIBUTION

Suppose $f_1, f_2, f_3, \ldots, f_n$ are the respective frequencies of the items $X_1, X_2, X_3, \ldots, X_n$. First we calculate the cumulative frequencies and then we see the median number $\left(\frac{n+1}{2}\right)$ under the cumulative frequency column. The item which corresponds to the median number is called median.

Example 3.19.

Given below is the frequency distribution of number of chairs in different rooms of a college. Find the median.

Number of chairs	30	33	40	45	52
Number of rooms	3	10	8	5	3

Solution:

The necessary calculations are given below:

Number of chairs (X)	Number of rooms (f)	Cumulative frequencies (c.f.)
30	3	3
33	10	3 + 10 = 13
40	8	13 + 8 = 21
45	5	21 + 5 = 26
52	3	26 + 3 = 29

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{29+1}{2}\right)\text{th item}$$

$$= \text{Value of (15) th item}$$

The cumulative frequency 15 corresponds to 40. Thus, Median = 40.

Example 3.20.

Given below is the frequency distribution of number of persons in 50 families in a village. Find the median as the average family size.

Number of persons	2	3	4	5	6
Number of families	5	8	12	20	5

Solution:

The necessary calculations are given below:

Number of persons (X)	Number of families (f)	Cumulative frequencies (c.f.)
2	5	5
3	8	5 + 8 = 13
4	12	13 + 12 = 25
5	20	25 + 20 = 45
6	5	45 + 5 = 50

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{50+1}{2}\right)\text{th item}$$

$$= \text{Value of (25.5) th item} = 5$$

Hence the average size of the family is 5 persons on the basis of median as an average.

3.6.2. MEDIAN FOR GROUPED DATA

For grouped data, we find the cumulative frequencies and then we calculate the median number $\frac{n}{2}$. The group which corresponds to the median number is called the median group. The median lies in this group. To interpolate median we use the formula

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - c\right)$$

where l = lower limit of the median class

h = size of the class interval of median class

f = frequency of the median class

n = sum of the frequencies

c = cumulative frequency of the class preceding the median class.

This formula is based on the assumption that the frequency (f) is uniformly spread in the median class. It is called linear interpolation of the median.

Example 3.21.

Find median from the following grouped data regarding heights of students in a college.

Heights (in inches)	56 – 58	58 – 60	60 – 62	62 – 64	64 – 66	66 – 68
Number of students	25	40	250	130	60	20

Solution:

The necessary calculations are given below:

Heights (in inches)	Number of students (f)	Cumulative frequencies (c.f.)
56 – 58	25	25
58 – 60	40	25 + 40 = 65
60 – 62	250	65 + 250 = 315
62 – 64	130	315 + 130 = 445
64 – 66	60	445 + 60 = 505
66 – 68	20	505 + 20 = 525

Median = Value of $\left(\frac{n}{2}\right)$ th item = Value of $\left(\frac{525}{2}\right)$ th item = 262.5th item

Median lies in the class 60 – 62.

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - c\right)$$

where $l = 60, h = 2, f = 250, \frac{n}{2} = 262.5$ and $c = 65$

Thus, Median $= 60 + \frac{2}{250}(262.5 - 65) = 60 + \frac{2}{250}(197.5) = 60 + 1.58 = 61.58$

Example 3.22.

Calculate median from the following data:

Groups	10 – 14	15 – 19	20 – 24	25 – 29	30 – 34
Frequency	5	12	30	25	6

Solution:

The necessary calculations are given below:

Groups	Class Boundaries (C.B)	f	Cumulative frequencies (c.f.)
10 – 14	9.5 – 14.5	5	5
15 – 19	14.5 – 19.5	12	5 + 12 = 17
20 – 24	19.5 – 24.5	30	17 + 30 = 47
25 – 29	24.5 – 29.5	25	47 + 25 = 72
30 – 34	29.5 – 34.5	6	72 + 6 = 78

Median = Value of $\left(\frac{n}{2}\right)$ th item = Value of $\left(\frac{78}{2}\right)$ th item = 39th item

Median lies in the class 19.5 – 24.5.

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - c\right)$$

where $l = 19.5,\ h = 5, f = 30,\ \frac{n}{2} = 39$ and $c = 17$

Thus, Median $= 19.5 + \frac{5}{30}(39 - 17) = 19.5 + \frac{5}{30}(22) = 19.5 + 3.67 = 23.17$

3.6.3. MERITS AND DEMERITS OF MEDIAN

MERITS

(i) It is easy to understand.

(ii) It is readily calculated.

(iii) Its graphic location is easy.

(iv) It is not affected by extreme values.

(v) It gives good results in a study of qualitative measurements.

(vi) It is an appropriate average in a highly skewed distribution.

(vii) Its value is least affected by the addition of a few more items.

(viii) It can be calculated even if the data has open intervals at either end provided the median does not lie in the open intervals.

DEMERITS

(i) It is not capable of further mathematical treatment. For example if we know the medians of two sets of data, we cannot find the combined median for both the sets.

(ii) It cannot give correct total when multiplied by the number of items.

(iii) It is necessary to arrange the values in an array before finding the median.

(iv) Median is more likely to be affected by the fluctuations of sampling than the arithmetic mean.

(v) If big or small items in a series are to receive greater importance median would be an unsuitable average. Median ignores the values of extreme items.

3.6.4. OTHER PARTITION VALUES

Median divides the observations into two equal parts. Thus 50 % observations are less than median and 50 % are above the median. There are certain other partition values which divide the observations into different groups. These are quartiles, deciles and percentiles.

3.6.5. QUARTILES

There are three quartiles called first quartile, second quartile and third quartile. These quartiles divide the set of observations into four equal parts. The second quartile is equal to the median. The first quartile is also called lower quartile and is denoted by Q_1. The third quartile is also called upper quartile and is denoted by Q_3. The lower quartile Q_1 is a point which has 25 % observations less than it and 75 % observations are above it. The upper quartile Q_3 is a point with 75 % observations below it and 25 % observations above it.

3.6.6. QUARTILES FOR INDIVIDUAL OBSERVATIONS (UNGROUPED DATA)

$$Q_1 = \text{Value of} \left(\frac{n+1}{4}\right) \text{th item}$$

$$Q_2 = \text{Value of } \frac{2(n+1)}{4} \text{ th item} \quad = \text{Value of} \left(\frac{n+1}{2}\right) \text{th item} = \text{Median}$$

$$Q_3 = \text{Value of } \frac{3(n+1)}{4} \text{ th item}$$

3.6.7. QUARTILES FOR A FREQUENCY DISTRIBUTION (DISCRETE DATA)

$$Q_1 = \text{Value of} \left(\frac{n+1}{4}\right) \text{th item} \qquad (n = \Sigma f)$$

$$Q_2 = \text{Value of } \frac{2(n+1)}{4} \text{ th item} = \text{Value of} \left(\frac{n+1}{2}\right) \text{th item} = \text{Median}$$

$$Q_3 = \text{Value of } \frac{3(n+1)}{4} \text{ th item}$$

3.6.8. QUARTILES FOR GROUPED FREQUENCY DISTRIBUTION

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right) \qquad (n = \Sigma f)$$

$$Q_2 = l + \frac{h}{f}\left(\frac{2n}{4} - c\right) \quad = l + \frac{h}{f}\left(\frac{n}{2} - c\right) = \text{Median}$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

3.6.9. DECILES

The deciles are the partition values which divide the set of observations into ten equal parts. There are nine deciles namely $D_1, D_2, D_3, \ldots, D_9$. The first decile D_1 is a point which has 10% of the observations below it.

3.6.10. DECILES FOR INDIVIDUAL OBSERVATIONS (UNGROUPED DATA)

$$D_1 = \text{Value of} \left(\frac{n+1}{10}\right) \text{th item}$$

$$D_2 = \text{Value of } \frac{2(n+1)}{10} \text{ th item}$$

$$D_3 = \text{Value of } \frac{3(n+1)}{10} \text{ th item}$$

$$\vdots$$

$$D_9 = \text{Value of } \frac{9(n+1)}{10} \text{ th item}$$

3.6.11. DECILES FOR A FREQUENCY DISTRIBUTION (DISCRETE DATA)

$$D_1 = \text{Value of } \left(\frac{n+1}{10}\right) \text{th item} \qquad (n = \Sigma f)$$

$$D_2 = \text{Value of } \frac{2(n+1)}{10} \text{ th item}$$

$$D_3 = \text{Value of } \frac{3(n+1)}{10} \text{ th item}$$

$$\vdots$$

$$D_9 = \text{Value of } \frac{9(n+1)}{10} \text{ th item}$$

3.6.12. DECILES FOR GROUPED FREQUENCY DISTRIBUTION

$$D_1 = l + \frac{h}{f}\left(\frac{n}{10} - c\right) \qquad (n = \Sigma f)$$

$$D_2 = l + \frac{h}{f}\left(\frac{2n}{10} - c\right)$$

$$D_3 = l + \frac{h}{f}\left(\frac{3n}{10} - c\right)$$

$$\vdots$$

$$D_9 = l + \frac{h}{f}\left(\frac{9n}{10} - c\right)$$

3.6.13. PERCENTILES

The percentiles are the points which divide the set of observations into one hundred equal parts. These points are denoted by $P_1, P_2, P_3, \ldots, P_{99}$, and are called the first, second, third, ..., ninty ninth percentiles. The percentiles are calculated for very large number of observations like workers in factories and the population in provinces or countries. The percentiles are usually calculated for grouped data. The first percentile denoted by P_1 is calculated as $P_1 = \text{Value of } \left(\frac{n}{100}\right)$ th item. We find the group in which the $\frac{n}{100}$ th item lies and then P_1 is interpolated from the formula

$$P_1 = l + \frac{h}{f}\left(\frac{n}{100} - c\right) \qquad P_2 = l + \frac{h}{f}\left(\frac{2n}{100} - c\right)$$

$$P_3 = l + \frac{h}{f}\left(\frac{3n}{100} - c\right) \ \ldots \qquad P_{99} = l + \frac{h}{f}\left(\frac{99n}{100} - c\right) \qquad (n = \Sigma f)$$

It may be noted that all the partition points can be expressed in terms of percentiles. For example, Median $= D_5 = P_{50}$, $Q_1 = P_{25}$, $Q_3 = P_{75}$, $D_1 = P_{10}$, $D_3 = P_{30}$, $D_7 = P_{70}$ and $D_9 = P_{90}$.

Example 3.23.

(a) Find lower quartile, median and upper quartile from the following data:

3, 7, 12, 25, 37, 48, 15, 69, 52, 73, 70, 88, 82, 80, 92.

(b) Compute lower and upper quartiles, 5th decile and 25th percentile from the given data:

26, 22, 14, 30, 18, 11, 35, 41, 12, 32.

Solution:

(a) After arranging the observations in ascending order. We have

3, 7, 12, 15, 25, 37, 48, 52, 69, 70, 73, 80, 82, 88, 92

$$Q_1 = \text{Value of}\left(\frac{n+1}{4}\right)\text{th item} = \text{Value of}\left(\frac{15+1}{4}\right)\text{th item}$$
$$= \text{Value of }(4)\text{ th item} = 15$$

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{15+1}{2}\right)\text{th item}$$
$$= \text{Value of }(8)\text{ th item} = 52$$

$$Q_3 = \text{Value of }\frac{3(n+1)}{4}\text{ th item} = \text{Value of }\frac{3(15+1)}{4}\text{ th item}$$
$$= \text{Value of }(12)\text{ th item} = 80$$

(b) After arranging the observations in ascending order, we get

11, 12, 14, 18, 22, 26, 30, 32, 35, 41

$$Q_1 = \text{Value of}\left(\frac{n+1}{4}\right)\text{th item} = \text{Value of}\left(\frac{10+1}{4}\right)\text{th item}$$
$$= \text{Value of }(2.75)\text{ th item}$$
$$= \text{2nd item} + 0.75\,(\text{3rd item} - \text{2nd item}) = 12 + 0.75\,(14 - 12)$$
$$= 12 + 1.5 = 13.5$$

$$Q_3 = \text{Value of }\frac{3(n+1)}{4}\text{ th item} = \text{Value of }\frac{3(10+1)}{4}\text{ th item}$$
$$= \text{Value of }(8.25)\text{ th item}$$
$$= \text{8th item} + 0.25\,(\text{9th item} - \text{8th item}) = 32 + 0.25\,(35 - 32)$$
$$= 32 + 0.75 = 32.75$$

$$D_5 = \text{Value of }\frac{5(n+1)}{10}\text{ th item} = \text{Value of }\frac{5(10+1)}{10}\text{ th item}$$
$$= \text{Value of }(5.5)\text{ th item}$$
$$= \text{5th item} + 0.5\,(\text{6th item} - \text{5th item}) = 22 + 0.5(26 - 22) = 22 + 2 = 24$$

$$P_{25} = \text{Value of }\frac{25(n+1)}{100}\text{ th item} = \text{Value of }\frac{25(10+1)}{100}\text{ th item}$$
$$= \text{Value of }(2.75)\text{ th item}$$
$$= \text{2nd item} + 0.75\,(\text{3rd item} - \text{2nd item}) = 12 + 0.75\,(14 - 12)$$
$$= 12 + 1.5 = 13.5$$

Example 3.24.

Eight coins were tossed together and the number of heads resulting was noted. The operation was repeated 256 times and the frequencies that were obtained for different values of X, the number of heads, are shown in the following table. Calculate median, quartiles, 6th decile and 27th percentile.

X	0	1	2	3	4	5	6	7	8
f	1	9	26	59	72	52	29	7	1

Solution:

Here the variable X is discrete and therefore X is written as 0, 1, 2, 3, ···, 8. The necessary calculations are given below:

X	0	1	2	3	4	5	6	7	8
f	1	9	26	59	72	52	29	7	1
c.f.	1	10	36	95	167	219	248	255	256

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{256+1}{2}\right)\text{th item}$$

$$= \text{Value of (128.5) th item} = 4$$

$$Q_1 = \text{Value of}\left(\frac{n+1}{4}\right)\text{th item} = \text{Value of}\left(\frac{256+1}{4}\right)\text{th item}$$

$$= \text{Value of (64.25) th item} = 3$$

$$Q_3 = \text{Value of } \frac{3(n+1)}{4}\text{ th item} = \text{Value of } \frac{3(256+1)}{4}\text{ th item}$$

$$= \text{Value of (192.75) th item} = 5$$

$$D_6 = \text{Value of } \frac{6(n+1)}{10}\text{ th item} = \text{Value of } \frac{6(256+1)}{10}\text{ th item}$$

$$= \text{Value of (154.2) th item} = 4$$

$$P_{27} = \text{Value of } \frac{27(n+1)}{100}\text{ th item} = \text{Value of } \frac{27(256+1)}{100}\text{ th item}$$

$$= \text{Value of (69.39) th item} = 3$$

Example 3.25.

Find median, lower and upper quartiles, 5th decile and 44th percentile from the following observations.

X	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2
f	6	9	15	24	33	20	13	7	5

Solution:

Here the variable X is continuous and the given values of X are the midpoints of the classes of grouped data. Hence we write the class boundaries. After arranging the observations in ascending order, we have

X	Class Boundaries	f	c.f.
0.2	0.1 – 0.3	5	5
0.4	0.3 – 0.5	7	5 + 7 = 12
0.6	0.5 – 0.7	13	12 + 13 = 25
0.8	0.7 – 0.9	20	25 + 20 = 45
1.0	0.9 – 1.1	33	45 + 33 = 78
1.2	1.1 – 1.3	24	78 + 24 = 102
1.4	1.3 – 1.5	15	102 + 15 = 117
1.6	1.5 – 1.7	9	117 + 9 = 126
1.8	1.7 – 1.9	6	126 + 6 = 132

Median = Value of $\left(\frac{n}{2}\right)$ th item = Value of $\left(\frac{132}{2}\right)$ th item = 66th item

Median lies in the class 0.9 – 1.1. Median $= l + \frac{h}{f}\left(\frac{n}{2} - c\right)$

where $l = 0.9$, $h = 0.2$, $f = 33$, $\frac{n}{2} = 66$ and $c = 45$

Thus, Median $= 0.9 + \frac{0.2}{33}(66 - 45) = 0.9 + \frac{0.2}{33}(21) = 0.9 + 0.127 = 1.027$

Q_1 = Value of $\left(\frac{n}{4}\right)$ th item = Value of $\left(\frac{132}{4}\right)$ th item = 33rd item

Q_1 lies in the class 0.7 – 0.9. $Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$

where $l = 0.7$, $h = 0.2$, $f = 20$, $\frac{n}{4} = 33$ and $c = 25$

Thus, $Q_1 = 0.7 + \frac{0.2}{20}(33 - 25) = 0.7 + \frac{0.2}{20}(8) = 0.7 + 0.08 = 0.78$

Q_3 = Value of $\left(\frac{3n}{4}\right)$ th item = Value of $\frac{3(132)}{4}$ th item = 99th item

Q_3 lies in the class 1.1 – 1.3. $Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$

Where $l = 1.1$, $h = 0.2$, $f = 24$, $\frac{3n}{4} = 99$ and $c = 78$

Thus, $Q_3 = 1.1 + \frac{0.2}{24}(99 - 78) = 1.1 + \frac{0.2}{24}(21) = 1.1 + 0.175 = 1.275$

$$D_5 = \text{Value of}\left(\frac{5n}{10}\right)\text{th item} = \text{Value of } \frac{5(132)}{10}\text{ th item} = 66\text{th item}$$

D_5 lies in the class $0.9 - 1.1$. $\quad D_5 = l + \frac{h}{f}\left(\frac{5n}{10} - c\right)$

Where, $l = 0.9, h = 0.2, f = 33, \frac{5n}{10} = 66$ and $c = 45$

Thus, $D_5 = 0.9 + \frac{0.2}{33}(66 - 45) = 0.9 + \frac{0.2}{33}(21) = 0.9 + 0.127 = 1.027$

$$P_{44} = \text{Value of}\left(\frac{44n}{100}\right)\text{th item} = \text{Value of } \frac{44(132)}{100}\text{ th item} = 58.08\text{th item}$$

P_{44} lies in the class $0.9 - 1.1$. $\quad P_{44} = l + \frac{h}{f}\left(\frac{44n}{100} - c\right)$

Where $l = 0.9$, $h = 0.2$, $f = 33$, $\frac{44n}{100} = 58.08$ and $c = 45$

Thus, $P_{44} = 0.9 + \frac{0.2}{33}(58.08 - 45) = 0.9 + \frac{0.2}{33}(13.08) = 0.9 + 0.079 = 0.979$

3.6.14. GRAPHIC LOCATION OF MEDIAN AND PARTITION VALUES

Median and other partition values can be located from the graph of the cumulative frequency polygon (ogive polygon) . Suppose we have a graph of the cumulative frequency polygon as shown in the diagram below :

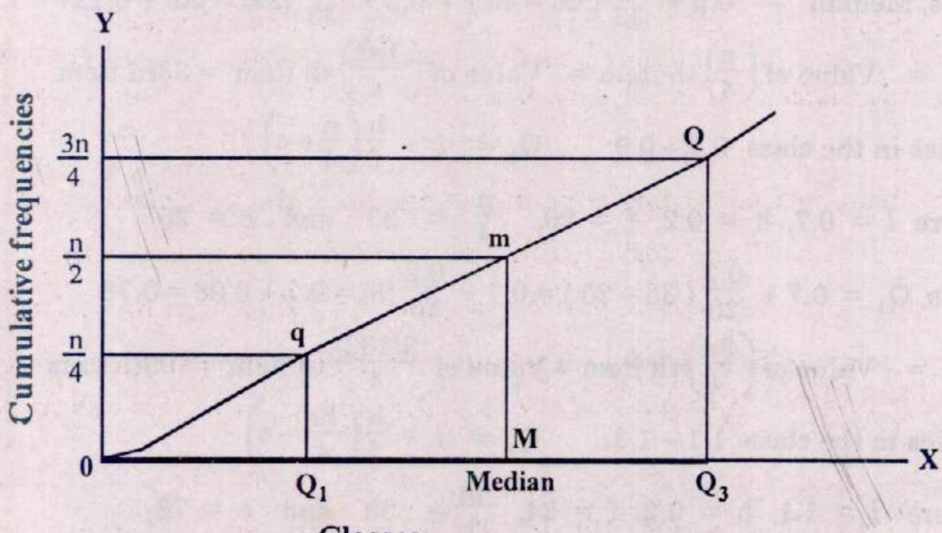


Fig.3.1. Graphical location of the partition values.

For median, we calculate n/2. On Y-axis, we mark the height equal to n/2 and from this point we draw a straight line parallel to X-axis which intersects the polygon at the point m. From the point m, we draw a perpendicular which touches the X-axis at M. This point on X-axis is the median. Similarly, for the lower quartile we take a height equal to n/4 on the Y-axis. From this point we draw a line parallel to X-axis which meets the polygon at the point q. From this point we draw a perpendicular on X-axis which touches it at the point Q_1 which is the first quartile. For upper quartile take the height on Y-axis equal to 3n/4 and proceed in a similar way as described above to find Q_3 and also the other partition values.

3.7. MODE

Mode is that observation which occurs maximum number of times in the data. In some data the mode may not exist, and even if it does exist it may not be unique. If there are two modes, in a distribution, the distribution is called a bimodal distribution.

Example 3.26.

Identify the mode for each of the following lists of numbers:

(i) 4, 2, 6, 6, 6, 5, 9, 10, 13 and 15

(ii) 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 and 9

(iii) 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, and 22

(iv) 8, 11, 4, 3, 2, 5, 10, 6, 4, 1, 10, 8, 12, 6, 5 and 7.

Solution:

(i) Since 6 occurs maximum number of times, therefore the mode = 6.

(ii) There are two modes which are 4 and 7 because both items occur maximum but equal number of times. A distribution with two modes is called a bimodal distribution.

(iii) Here, we have no mode because each item occurs the same number of times.

(iv) Since each of the numbers 4, 5, 6, 8 and 10 occurs twice, we can consider that there are five modes. However, it is more reasonable to conclude in this case that mode does not exist.

3.7.1. MODE FOR A FREQUENCY DISTRIBUTION (DISCRETE DATA)

In a discrete series the value of the variable against which the frequency is maximum would be the modal value.

Example 3.27.

The following are the size of shoes as worn by 10 persons. Calculate the modal size.

Size of shoes	4.0	4.5	5.0	5.5	6.0
Frequency	2	4	2	1	1

Solution:

Since the size 4.5 occurs the maximum number of times, therefore Mode = 4.5

3.7.2. MODE FOR GROUPED DATA

In case of grouped data, the mode lies in the class which has the maximum frequency. This class is called the modal class. This is done when all the class intervals are equal. If the class intervals are not uniform, the frequencies are converted into the frequencies per unit interval and then the class with maximum frequency is used as the modal class. The mode lies within the modal class. The following formula is used for locating the mode

$$\text{Mode} = l + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h = l + \frac{(f_m - f_1)}{2f_m - f_1 - f_2} \times h$$

where l = lower limit of the modal class

f_m = maximum frequency of the modal class

f_1 = frequency of the class preceding the modal class

f_2 = frequency of the class following the modal class

h = length of the class interval of modal class.

Note 1: Sometimes two classes have equal maximum frequencies, in that case mode should not be calculated.

Note 2: While calculating the value of mode it should be seen that the class intervals of the different classes are equal otherwise the above formula cannot be applied. When mode is illdefined the above formula is not applicable. In that case the value of mode is indirectly obtained by applying the following formula:

Mode = 3 Median – 2 Mean.

Example 3.28.

Compute the mode of the following distribution:

Classes	0 – 7	7 – 14	14 – 21	21 – 28	28 – 35	35 – 42	42 – 49	49 – 56	56 – 63
Frequency	3	11	15	20	25	18	13	3	2

Solution:

Here the maximum frequency is 25 belonging to the class 28 – 35. So the modal class is 28 – 35. Using the formula,

$$\text{Mode} = l + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where $l = 28$, $f_m = 25$, $f_1 = 20$, $f_2 = 18$ and $h = 7$

$$\text{Thus, Mode} = 28 + \frac{(25 - 20)}{(25 - 20) + (25 - 18)} \times 7 = 28 + \frac{5}{5 + 7} \times 7 = 28 + 2.917 = 30.917$$

3.7.3. MERITS AND DEMERITS OF MODE

MERITS

(i) It is easily located.

(ii) It is not affected by the extreme values.

(iii) It can be located in open ended classes.

(iv) It is the most representative of the averages.

(v) It is easily understood by a common man.

DEMERITS

(i) It has no significance when number of items is not large.

(ii) It is not based on all the observations.

(iii) It is not capable of further mathematical treatment.

(iv) Sometimes a distribution may have more than one mode. In this case mode should not be calculated.

3.8. EMPIRICAL RELATION BETWEEN MEAN, MEDIAN AND MODE

In a moderately skewed distribution, median lies in between the other two averages i.e. mean and mode. For a moderately skewed distribution there exists an empirical relationship among the mean, median and mode. The difference between mean and median is half of the distance between median and mode. The difference between mean and mode is three times the difference between mean and median. The relation between them is

$$\text{Mean} - \text{Median} = \frac{1}{2}(\text{Median} - \text{Mode})$$

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$$

$$\text{Median} - \text{Mode} = 2\,\text{Mean} - 2\,\text{Median} = 2(\text{Mean} - \text{Median})$$

$$\text{Mean} = \frac{1}{2}(3\,\text{Median} - \text{Mode})$$

$$\text{Median} = \frac{1}{3}(2\,\text{Mean} + \text{Mode})$$

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

Example 3.29.

(i) In a moderately skewed distribution, arithmetic mean = 24.6 and the mode = 26.1. Find the value of the median and explain the reason for the method employed.

(ii) In moderately asymmetrical distribution, the value of median is 42.8 and the value of mode is 40. Find the mean.

(iii) Find out the missing figure.
Mean = ? (3 Median – Mode)

(iv) In a moderately asymmetrical distribution, the value of mean and median is 15 and 16 respectively. Find out the value of mode.

Solution:

The necessary calculations are given below:

(i) In a moderately asymmetrical distribution the following relationship holds good amongst mean, median and mode.

$$\text{Median} = \frac{1}{3}(2\,\text{Mean} + \text{Mode})$$

Hence given any two averages, we can determine the third.

Here, Mean = 24.6 and Mode = 26.1

$$\text{Thus, Median} = \frac{1}{3}(2 \times 24.6 + 26.1) = \frac{1}{3}(49.2 + 26.1) = 25.1$$

(ii) Mean $= \frac{1}{2}$ (3 Median – Mode)

Here, Median = 42.8 and Mode = 40

Thus, Mean $= \frac{1}{2}(3 \times 42.8 - 40) = \frac{1}{2}(128.4 - 40) = 44.2$

(iii) Mean $= \frac{1}{2}$ (3 Median – Mode). Hence the missing figure is $\frac{1}{2}$.

(iv) Mode = 3 Median – 2 Mean

Here Mean = 15 and Median = 16. Thus, Mode = 3(16) – 2(15) = 48 – 30 = 18

3.9. GEOMETRIC MEAN

The geometric mean 'G' of n positive values is defined as the nth root of their product. Thus, it is obtained by multiplying together all the n values and then taking the nth root of the product. Thus,

$$G = [X_1 . X_2 . X_3 . \dots . X_n]^{\frac{1}{n}} = \sqrt[n]{X_1 . X_2 . X_3 . \dots . X_n}$$

Where n stands for the number of observations and $X_1, X_2, X_3, \dots, X_n$ are the various values. For instance the geometric mean 'G' of 2, 4, 8 is

$$G = [2 \times 4 \times 8]^{\frac{1}{3}} = [64]^{\frac{1}{3}} = 4$$

The above method of calculating geometric mean is satisfactory only if there are two or three observations. But if n is a large number, the problem of computing the nth root of the product of these values by simple arithmetic is a tedious work. To facilitate the computation of geometric mean we make use of logarithms. The above formula when reduced to its logarithmic form can be written as :

$$\log G = \frac{1}{n}[\log X_1 + \log X_2 + \log X_3 + \cdots + \log X_n] = \frac{1}{n}\Sigma \log X_i$$

$$\text{or} \qquad G = \text{Antilog}\left(\frac{\Sigma \log X_i}{n}\right) \qquad \text{(For ungrouped data)}$$

The logarithm of the geometric mean is equal to the arithmetic mean of the logarithms of individual values. The geometric mean is used when we are interested in averaging the ratios, for example, in the construction of index numbers, the geometric mean is the most suitable average for averaging the price or link relatives.

If the n non-zero and positive variate-values $X_1, X_2, X_3, \dots, X_n$ occur $f_1, f_2, f_3, \dots, f_n$ times respectively, then geometric mean G, is defined by

$$G = \left[X_1^{f_1} \cdot X_2^{f_2} \cdot X_3^{f_3} \cdot \dots \cdot X_n^{f_n}\right]^{\frac{1}{N}} \quad \text{where } N = f_1 + f_2 + f_3 + \dots + f_n = \Sigma f_i$$

$$\log G = \frac{1}{N}[f_1 \log X_1 + f_2 \log X_2 + f_3 \log X_3 + \dots + f_n \log X_n] = \frac{1}{N}\Sigma f_i \log X_i$$

$$G = \text{Antilog}\left(\frac{\Sigma f_i \log X_i}{N}\right) \text{ or } G = \text{Antilog}\left(\frac{\Sigma f_i \log X_i}{\Sigma f_i}\right) \quad \text{(For grouped data)}$$

Example 3.30

Given the following set of data from a sample of size 6:

6, 3, 8, 4, 2, 10

Calculate the geometric mean.

Solution:

The necessary calculations are given below:

X	6	3	8	4	2	10	Total
log X	0.7782	0.4771	0.9031	0.6021	0.3010	1.0000	4.0615

$$G = \text{Antilog}\left(\frac{\Sigma \log X}{n}\right) = \text{Antilog}\left(\frac{4.0615}{6}\right) = 4.7524$$

Example 3.31.

The logarithms of 10 values of X are: 1.0792, 1.1761, 0.9542, 1.1761, 1.2041, 1.2553, 1.3010, 1.3979, 1.3010, 1.4771. Calculate the geometric mean of X values.

Solution:

Here, $n = 10$ and $\Sigma \log X_i = 12.3220$. Therefore

$$G = \text{Antilog}\left(\frac{\Sigma \log X_i}{n}\right) = \text{Antilog}\left(\frac{12.3220}{10}\right)$$

$$= \text{Antilog}(1.2322) = 17.0687$$

Example 3.32.

The following are the reciprocals of the values of the variable X. Find the geometric mean of X: 0.1345, 0.0378, 1.2937, 0.4588, 0.0039.

Solution:

The necessary calculations are given below:

1/X	X	log X
0.1345	7.43	0.8710
0.0378	26.46	1.4226
1.2937	0.77	−0.1135
0.4588	2.18	0.3385
0.0039	256.41	2.4089
Total	---	4.9275

Here, $n = 5$, $\Sigma \log X = 4.9275$

$$G = \text{Antilog}\left(\frac{\Sigma \log X}{n}\right)$$

$$= \text{Antilog}\left(\frac{4.9275}{5}\right)$$

$$= \text{Antilog}(0.9855) = 9.67$$

Example 3.33.

Find the average rate of increase in population which in the first decade increased by 20 %, in the next by 30 % and in the third by 40 %.

Solution:

The necessary calculations are given below:

Decade	Population (X)	log X
I ..	100 + 20 = 120	2.0792
II	100 + 30 = 130	2.1139
III	100 + 40 = 140	2.1461
Total	---	6.3392

$$G = \text{Antilog}\left(\frac{\Sigma \log X}{n}\right) = \text{Antilog}\left(\frac{6.3392}{3}\right) = \text{Antilog }(2.1131) = 129.75$$

Average rate of increase = 129.75 – 100 = 29.75 %.

Example 3.34.

Find the two numbers whose arithmetic mean is 9 and their geometric mean is 7.2.

Solution:

Let the two numbers be X and Y so that

$$\frac{X+Y}{2} = 9 \quad \text{and} \quad \sqrt{XY} = 7.2 \quad \text{or} \quad XY = (7.2)^2 = 51.84$$

$$X + Y = 18 \quad \text{and} \quad XY = 51.84$$

We know that $(X-Y)^2 = (X+Y)^2 - 4XY$

Substituting the values, we get

$$(X-Y)^2 = (18)^2 - 4(51.84) = 324 - 207.36 = 116.64$$

$$X - Y = \sqrt{116.64} = 10.8 \quad \cdots\cdots \quad (1)$$

$$X + Y = 18 \quad \cdots\cdots \quad (2)$$

Solving equations (1) and (2) ,we get

$$2X = 28.8 \quad \text{or} \quad X = 14.4$$

Substituting $X = 14.4$ in equation (2) , we get

$$14.4 + Y = 18 \text{ or } Y = 18 - 14.4 = 3.6$$

Thus, the two numbers are 14.4 and 3.6.

Example 3.35.

Calculate geometric mean for the following distribution.

Weights (in lbs.)	100 – 104	105 – 109	110 – 114	115 – 119	120 – 124
Frequency	24	30	45	65	72

Solution:

The necessary calculations are given below:

Weights (in lbs.)	X	f	log X	f log X
100 – 104	102	24	2.0086	48.2064
105 – 109	107	30	2.0294	60.8820
110 – 114	112	45	2.0492	92.2140
115 – 119	117	65	2.0682	134.4330
120 – 124	122	72	2.0864	150.2208
Total	---	236	----	485.9562

$$G = \text{Antilog}\left(\frac{\Sigma f \log X}{\Sigma f}\right) = \text{Antilog}\left(\frac{485.9562}{236}\right) = \text{Antilog}\,(2.0591) = 114.58$$

3.9.1. PROPERTIES OF GEOMETRIC MEAN

The main properties of geometric mean are:

(1) The geometric mean is less than arithmetic mean, i.e. G.M. < A.M.

(2) The product of the items remains unchanged if each item is replaced by the geometric mean.

(3) The geometric mean of the ratio of corresponding observations in two series is equal to the ratios of their geometric means.

(4) The geometric mean of the products of corresponding items in two series is equal to the product of their geometric means.

3.9.2. MERITS AND DEMERITS OF GEOMETRIC MEAN

MERITS

(i) It is rigidly defined and its value is a precise figure.

(ii) It is based on all the observations.

(iii) It is capable of further algebraic treatment.

(iv) It is not much affected by fluctuations of sampling.

(v) It is not much affected by extreme values.

DEMERITS

(i) It cannot be calculated if any of the observations is zero or negative.

(ii) Its calculation is rather difficult.

(iii) It is not easy to understand.

(iv) It may not coincide with any of the observations.

3.10. HARMONIC MEAN

Harmonic mean is defined as the reciprocal of the mean of the reciprocals of the items in a series. It is the ratio of the number of items and the sum of reciprocals of items. If the observations are $X_1, X_2, X_3 \cdots, X_n$, then the harmonic mean H is

$$H = \frac{\text{Number of items}}{\text{Sum of reciprocals of items}} = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \cdots + \frac{1}{X_n}} = \frac{n}{\Sigma\left(\frac{1}{X_i}\right)}$$

The calculation of H is not possible if any item is equal to zero.

If the midpoints of the classes are $X_1, X_2, X_3, \cdots, X_n$ with corresponding frequencies $f_1, f_2, f_3, \cdots, f_n$, then the harmonic mean H is given by

$$H = \frac{f_1 + f_2 + f_3 + \cdots + f_n}{\frac{f_1}{X_1} + \frac{f_2}{X_2} + \frac{f_3}{X_3} + \cdots \frac{f_n}{X_n}} = \frac{\Sigma f}{\Sigma\left(\frac{f}{X}\right)}$$

Example 3.36.

Given the following set of data from a sample of size 5:

13.2, 14.2 14.8 15.2, 16.1

Calculate the harmonic mean.

Solution:

The harmonic mean is calculated as below :

X	1/X
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\Sigma(1/X) = 0.3417$

$$H = \frac{n}{\Sigma\left(\frac{1}{X}\right)} = \frac{5}{0.3417} = 14.63$$

Example 3.37.

Calculate the harmonic mean for the data given below :

Marks	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89	90 – 99
Frequency	2	3	11	20	32	25	7

Solution. The necessary calculations are given below :

Marks	X	f	f/X
30 – 39	34.5	2	0.0580
40 – 49	44.5	3	0.0674
50 – 59	54.5	11	0.2018
60 – 69	64.5	20	0.3101
70 – 79	74.5	32	0.4295
80 – 89	84.5	25	0.2959
90 – 99	94.5	7	0.0741
Total	...	100	1.4368

$$H = \frac{\Sigma f}{\Sigma\left(\frac{f}{X}\right)} = \frac{100}{1.4368} = 69.60$$

Example 3.38.

The daily wages for a group of 236 persons have been obtained from a frequency distribution of a continuous variable X, after making the substitution $u = \frac{X - 112}{5}$.

$u = \frac{X-112}{5}$	– 2	– 1	0	1	2
No. of persons	24	30	45	65	72

Calculate the harmonic mean.

Solution:

The necessary calculations are given below:

$u = \frac{X-112}{5}$	X = 5u + 112	f	$\frac{f}{X}$
– 2	102	24	0.2353
– 1	107	30	0.2804
0	112	45	0.4018
1	117	65	0.5556
2	122	72	0.5902
Total	...	236	2.0633

$$H = \frac{\Sigma f}{\Sigma\left(\frac{f}{X}\right)} = \frac{236}{2.0633} = 114.38$$

3.10.1. MERITS AND DEMERITS OF HARMONIC MEAN

Merits

(i) It is based on all the observations.

(ii) It is not much affected by the fluctuations of sampling.

(iii) It is capable of algebraic treatment.

(iv) It is an appropriate average for averaging ratios and rates.

(v) It does not give much weight to the large items.

Demerits

(i) Its calculation is difficult.

(ii) It gives high weightage to the small items.

(iii) It cannot be calculated if any one of the items is zero.

(iv) It is usually a value which does not exist in the given data.

SHORT DEFINITIONS

Measure of Central Tendency

A measure of central tendency is defined as a single value that is considered most representative of the whole set of data.

or

A descriptive measure that indicates the central position in a set of data is called measure of central tendency.

Average

An average is a central value which can represent the whole data.

or

The average is a value which locates the central position of the observations.

Mean

The mean of a set of data is found by adding up all the observations and dividing by the total number of observations.

or

The mean of a set of quantitative data is equal to the sum of the measurements divided by the number of measurements contained in the data set. $\bar{x} = \frac{\Sigma x}{}$

Weighted Mean $\bar{x} = \frac{\Sigma w \bar{x}}{\Sigma w}$

The mean for a set of data obtained by assigning each data value a weight that reflects its relative importance within the set, is called weighted mean.

or

The mean of values when each value is weighted according to its relative importance, is called weighted mean.

Median The Middle Most value in an arrange Data is called Median

The median is the midpoint of the values after they have been ordered from the smallest to the largest or the largest to the smallest.

it is Donated by $\tilde{x}$ its formular as $\tilde{x} = \frac{n+1}{2}$ the value.

The median of a data is the middle item when the items are arranged in ascending or descending order.

Quartiles The middle most value in an arrange Data is called Quartiles

Quartiles are the values of the variate that divide a set of data into four equal parts after arranging the observations in ascending order of magnitude. These are Q_1 (lower quartile), Q_2 (Median) and Q_3 (upper quartile)

Deciles

Deciles are the values of the variate that divide a set of data into ten equal parts after arranging the observations in ascending order of magnitude. $D_1, D_2, \ldots, D_9$

Percentile

Percentiles are the values of the variate that divide a set of data into one hundred equal parts after arranging the observations in ascending order of magnitude. $P_1, P_2 \ldots P_{99}$

Mode → The Most Repeated ~~Data~~ Value in case Data is called Mode.

The mode is the measurement that occurs with greatest frequency in a set of data.

Mode Maybe More then 1. if there is no Repeate value in the Data then its Mode is "no Mode"

The value of the variable with the largest frequency is called the mode.

Symmetric Data

A set of data whose values are evenly spread around the centre is known as symmetric data. For symmetric data, the mean and the median are equal.

Skewed Data

A set of data that is not symmetrical is skewed data. For skewed data, the mean will be larger or smaller than the median.

Left-Skewed Data

A data distribution is left skewed if the mean for the data is smaller than the median.

Right-Skewed Data

A data distribution is right skewed if the mean for the data is larger than the median.

Geometric Mean The nth Root of the Product of the nth Positive Value is called Geometric Mean. or its formula $G.M = (x_1, x_2, x_3 \ldots x_n)^{1/n}$

The geometric mean of a set of n positive numbers is the nth root of their product.

The geometric mean of non-zero and positive values is obtained by multiplying all the values and then extracting the relevant root of the product.

Harmonic Mean →

Harmonic mean of a series is the reciprocal of the arithmetic mean of the reciprocals of the values

or

The harmonic mean is computed by averaging the reciprocals of the values of the variable and finding the reciprocal of this average.

Relative Position of Averages

(i) In a symmetrical distribution, the mean, median and mode coincide.

(ii) In a moderately asymmetrical distribution, the mean, median and mode are related as mean − mode = 3 (mean − median)

(iii) If all the values in a variable are the same, then A.M. = G.M. = H.M. But if the values vary and none of values is zero or negative, then A.M. > G.M. > H.M.

(iv) The geometric mean of any two values is equal to the geometric mean of their arithmetic mean and harmonic mean.

MULTIPLE – CHOICE QUESTIONS

1. Any measure indicating the centre of a set of data, arranged in an increasing or decreasing order of magnitude, is called a measure of:
   (a) skewness (b) symmetry
   (c) central tendency (d) dispersion
2. Scores that differ greatly from the measures of central tendency are called:
   (a) raw scores (b) the best scores
   (c) extreme scores (d) z-scores
3. The measure of central tendency listed below is:
   (a) the raw score (b) the mean
   (c) the range (d) standard deviation
4. The total of all the observations divided by the number of observations is called:
   (a) arithmetic mean (b) geometric mean
   (c) median (d) harmonic mean
5. While computing the arithmetic mean of a frequency distribution, the each value of a class is considered equal to:
   (a) class mark (b) lower limit
   (c) upper limit (d) lower class boundary
6. Change of origin and scale is used for calculation of the:
   (a) arithmetic mean (b) geometric mean
   (c) weighted mean (d) lower and upper quartiles
7. The sample mean $\bar{X}$ is a:
   (a) parameter (b) statistic
   (c) variable (d) constant
8. The population mean μ is called:
   (a) discrete variable (b) continuous variable
   (c) parameter (d) sampling unit
9. The arithmetic mean is highly affected by:
   (a) moderate values (b) extremely small values
   (c) odd values (d) extremely large values
10. The sample mean $\bar{X}$ is calculated by the formula:
   (a) $\frac{\sum fX}{\sum f}$ (b) $A + \frac{\sum fD}{\sum f}$
   (c) $A + \frac{\sum fu}{\sum f} \times h$ (d) all of the above
11. If a constant value is added to every observation of data, then arithmetic mean is obtained by:
   (a) subtracting the constant (b) adding the constant
   (c) multiplying the constant (d) dividing the constant

12. Which of the following statements is always true?

(a) The mean has an effect on extreme scores

(b) The median has an effect on extreme scores

(c) Extreme scores have an effect on the mean

(d) Extreme scores have an effect on the median

13. The elimination of extreme scores at the bottom of the set has the effect of:

(a) lowering the mean (b) raising the mean

(c) no effect (d) none of the above

14. The elimination of extreme scores at the top of the set has the effect of:

(a) lowering the mean (b) raising the mean

(c) no effect (d) difficult to tell

15. The sum of the deviations taken from mean is :

(a) always equal to zero. (b) some times equal to zero

(c) never equal to zero. (d) less than zero.

16. If $\bar{X} = 25$, which of the following will be minimum:

(a) $\Sigma(X-27)^2$ (b) $\Sigma(X-25)^2$

(c) $\Sigma(X-22)^2$. (d) $\Sigma(X+25)^2$

17. The sum of the squares of the deviations about mean is:

(a) zero (b) maximum

(c) minimum (d) all of the above

18. If $\sum_{i=1}^{10}(X_i - 50) = 100$, then sample mean $\bar{X}$ will be:

(a) 10 (b) 50

(c) 60 (d) 100

19. For a certain distribution, if $\Sigma(X-20) = 25$, $\Sigma(X-25) = 0$, and $\Sigma(X-35)$

then $\bar{X}$ is equal to:

(a) 20 (b) 25

(c) -25 (d) 35

20. The sum of the squares of the deviations of the values of a variable is least when the deviations are measured from:

(a) harmonic mean (b) geometric mean

(c) median (d) arithmetic mean

21. If $X_1, X_2, X_3, \ldots X_n$, be n observations having arithmetic mean $\bar{X}$ and if $Y = 4X \pm 2$, then $\bar{Y}$ is equal to:

(a) $4X$ (b) $4\bar{X}$

(c) $4\bar{X} \pm 2$ (d) 4 ± 2

22. If $\bar{X} = 100$ and $Y = 2X - 200$, then mean of Y values will be:

(a) 0 (b) 2

(c) 100 (d) 200

23. Step deviation method or coding method is used for computation of the:

(a) arithmetic mean (b) geometric mean

(c) weighted mean (d) harmonic mean

24. If the arithmetic mean of 20 values is 10, then sum of these 20 values is:

(a) 10 (b) 20

(c) 200 (d) 20 + 10

25. Ten families have an average of 2 boys. How many boys do they have together?

(a) 2 (b) 10

(c) 12 (d) 20

26. If the arithmetic mean of the two numbers X_1 and X_2 is 5 if $X_1 = 3$, then X_2 is:

(a) 3 (b) 5

(c) 7 (d) 10

27. Given $X_1 = 20$ and $X_2 = -20$. The arithmetic mean will be:

(a) zero (b) infinity

(c) impossible (d) difficult to tell

28. The mean of 10 observations is 10. All the observations are increased by 10 %. The mean of increased observations will be:

(a) 10 (b) 1.1

(c) 10.1 (d) 11

29. The frequency distribution of the hourly wage rate of 60 employees of a paper mill is as follows:

Wage Rate (Rs.)	54 – 56	56 – 58	58 – 60	60 – 62	62 – 64
Number of Workers	10	10	20	10	10

The mean wage rate is:

(a) Rs. 58.60 (b) Rs. 59.00

(c) Rs. 57.60 (d) Rs. 57.10

30. The sample mean $\bar{X}$ of first n natural numbers is:

(a) $n(n + 1)/2$ (b) $(n - 1)/2$

(c) $n/2$ (d) $(n + 1)/2$

31. The mean of first 2n natural numbers is:

(a) $\frac{n(n+1)}{2}$ (b) $\frac{n(2n+1)}{2}$

(c) $\frac{2n+1}{2}$ (d) $\frac{n+1}{2}$

32. The sum of deviations is zero when deviations are taken from:
(a) mean. (b) median.
(c) mode. (d) geometric mean

33. When the values in a series are not of equal importance, we calculate the:
(a) arithmetic mean (b) geometric mean
(c) weighted mean (d) mode

34. When all the values in a series occur the equal number of times, then it is not possible to calculate the:
(a) arithmetic mean (b) geometric mean
(c) harmonic mean (d) weighted mean

35. The mean for a set of data obtained by assigning each data value a weight that reflects its relative importance within the set, is called:
(a) geometric mean (b) harmonic mean
(c) weighted mean (d) combined mean

36. If $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k$ be the arithmetic means of **k** distributions with respective frequencies $n_1, n_2, n_3, \ldots, n_k$, then the mean of the whole distribution $\bar{X}_c$ is given by:
(a) $\Sigma\bar{X}/\Sigma n$ (b) $\Sigma n/\Sigma\bar{X}$
(c) $\Sigma n\bar{X}/\Sigma n$ (d) $\Sigma(n+\bar{X})/\Sigma n$

37. The combined arithmetic mean is calculated by the formula:
(a) $\dfrac{\bar{X}_1+\bar{X}_2}{n_1+n_2}$ (b) $\dfrac{n_1+n_2}{2}$
(c) $\dfrac{n_1\bar{X}_1+n_2\bar{X}_2}{n_1+n_2}$ (d) $\dfrac{n_1X_1+n_2X_2}{2}$

38. The arithmetic mean of 10 items is 4 and the arithmetic mean of 5 items is 10. The combined arithmetic mean is:
(a) 4 (b) 5
(c) 6 (d) 90

39. The midpoint of the values after they have been ordered from the smallest to the largest or the largest to the smallest is called:
(a) mean (b) median
(c) lower quartile (d) upper quartile

40. The first step in calculating the median of a discrete variable is to determine the:
(a) cumulative frequencies (b) relative weights
(c) relative frequencies (d) array

41. The suitable average for qualitative data is:
(a) mean (b) median
(c) mode (d) geometric mean

42. Extreme scores will have the following effect on the median of an examination:
(a) they may have no effect on it (b) they may tend to raise it
(c) they may tend to lower it (d) none of the above

43. We must arrange the data before calculating:
(a) mean. (b) median.
(c) mode. (d) geometric mean

44. If the smallest observation in a data is decreased, the average which is not affected is:
(a) mode. (b) median.
(c) mean. (d) harmonic mean.

45. If the data contains an extreme value, the suitable average is:
(a) mean. (b) median.
(c) weighted mean. (d) mode.

46. Sum of absolute deviations of the values is least when deviations are taken from:
(a) mean (b) mode
(c) median (d) Q_3

47. The frequency distribution of the hourly wage rate of 100 employees of a paper mill is as follows:

Wage Rate (Rs.)	54 – 56	56 – 58	58 – 60	60 – 62	62 – 64
Number of Workers	20	20	20	20	20

The median wage rate is:
(a) Rs. 55 (b) Rs. 57
(c) Rs. 56 (d) Rs. 59

48. The values of the variate that divide a set of data into four equal parts after arranging the observations in ascending order of magnitude are called:
(a) quartiles (b) deciles
(c) percentiles (d) difficult to tell

49. The lower and upper quartiles of a symmetrical distribution are 40 and 60 respectively. The value of median is:
(a) 40 (b) 50
(c) 60 (d) (60 – 40) /2

50. If in a discrete series 75% values are less than 30, then:
(a) $Q_3 < 75$ (b) $Q_3 < 30$
(c) $Q_3 = 30$ (d) $Q_3 > 30$

51. If in a discrete series 75% values are greater than 50, then:
(a) $Q_1 = 50$ (b) $Q_1 < 50$
(c) $Q_1 > 50$ (d) $Q_1 \neq 50$

52. If in a discrete series 25% values are greater than 75, then:
(a) $Q_1 > 75$ (b) $Q_1 = 75$
(c) $Q_3 = 75$ (d) $Q_3 > 75$

53. If in a discrete series 40% values are less than 40, then:
(a) $D_4 \neq 40$ (b) $D_4 < 40$
(c) $D_4 > 40$ (d) $D_4 = 40$

54. If in a discrete series 15% values are greater than 70, then
(a) $P_{15} = 70$ (b) $P_{85} = 15$
(c) $P_{85} = 70$ (d) $P_{70} = 70$

55. The middle value of an ordered series is called:
(a) median (b) 5th decile
(c) 50th percentile (d) all the above

56. If in a discrete series 50% values are less than 50, then:
(a) $Q_2 = 50$ (b) $D_5 = 50$
(c) $P_{50} = 50$ (d) all of the above

57. The mode or modal value of the distribution is that value of the variate for which frequency is:
(a) minimum (b) maximum
(c) odd number (d) even number

58. Suitable average for averaging the shoe sizes for children is:
(a) mean. (b) mode.
(c) median. (d) geometric mean.

59. Extreme scores on an examination have the following effect on the mode:
(a) they tend to raise it (b) they tend to lower it
(c) they have no effect on it (d) difficult to tell

60. A measurement that corresponds to largest frequency in a set of data is called:
(a) mean (b) median
(c) mode (d) percentile

61. Which of the following average cannot be calculated for the observations 2,2,4,4,6,6,8,8,10,10?
(a) mean (b) median
(c) mode (d) all of the above

62. Mode of the series 0,0,0,2,2,3,3,8,10 is:
(a) 0 (b) 2
(c) 3 (d) no mode

63. A distribution with two modes is called:
(a) unimodal (b) bimodal
(c) multimodal (d) normal

64. The modal letter of the word "STATISTICS" is:
(a) S (b) T
(c) both S and I (d) both S and T

65. The mode for the following frequency distribution is:

Weekly sales of burner units	0	1	2	3	Over 3
Number of weeks	38	6	5	1	0

(a) 50 (b) 38
(c) 0 (d) no mode

66. Which of the following statements is always correct:
(a) Mean = Median = Mode
(b) Arithmetic mean = Geometric mean = Harmonic mean
(c) Median = $Q_2 = D_5 = P_{50}$
(d) Mode = 2 Median −3 Mean

67. In a moderately asymmetrical series, the arithmetic mean, median and mode are related as:
(a) mean − mode = 3 (mean − median)
(b) mean − median = 2 (median − mode)
(c) median − mode = (mean − median)/2
(d) mode − median = 2 mean − 2 median

68. In a moderately skewed distribution, mean is equal to:
(a) (3median − mode)/2 (b) (2mean + mode)/3
(c) 3 median − 2 mean (d) 3 median − mode

69. In a moderately asymmetrical distribution, the value of median is given by:
(a) 3 median + 2 mean (b) 2 mean + mode
(c) (2 mean + mode)/3 (d) (3 median − mode)/2

70. For moderately skewed distribution, the value of mode is calculated as:
(a) 2 mean − 3 median (b) 3 median −2 mean
(c) 2 mean + mode (d) 3 median − mode

71. In a moderately skewed distribution, mean = 45 and median = 30, then the value of mode is:
(a) 0 (b) 30
(c) 45 (d) 180

72. If for any frequency distribution, the median is 10 and the mode is 30, then approximate value of mean is equal to:
(a) 0 (b) 10
(c) 30 (d) 60

73. In a moderately asymmetrical distribution, the value of mean and mode is 15 and 18 respectively. The value of median will be:

(a) 48 (b) 18
(c) 16 (d) 15

74. For moderately skewed distribution $\frac{\text{median} - \text{mode}}{\text{mean} - \text{median}}$ is equal to:

(a) 2 (b) 3
(c) 1/2 (d) 1/3

75. Which of the following is correct in a positively skewed distribution?

(a) Mean = Median = Mode (b) Mean < Median < Mode
(c) Mean > Median > Mode (d) Mean + Median + Mode

76. If the values of mean, median and mode coincide in a unimodal distribution, then the distribution will be:

(a) skewed to the left (b) skewed to the right
(c) multimodal (d) symmetrical

77. A curve that tails off to the right end, is called:

(a) positively skewed (b) negatively skewed
(c) symmetrical (d) both (b) and (c)

78. In a unimodal symmetrical distribution the highest point on the curve is called the:

(a) mean (b) median
(c) mode (d) all of the above

79. If a set of data has one mode and its value is less than mean, then the distribution is called:

(a) positively skewed (b) negatively skewed
(c) symmetrical (d) normal

80. Taking the relevant root of the product of all non-zero and positive values is called:

(a) arithmetic mean (b) geometric mean
(c) harmonic mean (d) combined mean

81. The best average in percentage rates and ratios is:

(a) arithmetic mean (b) lower and upper quartiles
(c) geometric mean (d) harmonic mean

82. The suitable average for computing average percentage increase in population is:

(a) geometric mean (b) harmonic mean
(c) combined mean (d) population mean

83. If 10% is added to each value of variable, the geometric mean of new variable is added by:

(a) 10 (b) 1/100
(c) 10% (d) 1.1

84. If each observation of a variable X is increased by 20%, then geometric mean is also increased by:
(a) 20 (b) 1/20
(c) 20% (d) 100%

85. If any value in a series is negative, then we cannot calculate the:
(a) mean (b) median
(c) geometric mean (d) harmonic mean

86. Geometric mean for X_1 and X_2 is:
(a) $\sqrt{X_1 + X_2}$ (b) $\sqrt{X_1X_2}$
(c) $\sqrt{X_1} + \sqrt{X_2}$ (d) $\sqrt{2X_1X_2}$

87. Geometric mean of 2, 4, 8 is:
(a) 6 (b) 4
(c) 14/3 (d) 8

88. Geometric mean is suitable when the values are given as:
(a) proportions (b) ratios
(c) percentage rates (d) all of the above

89. If the geometric mean of the two numbers X_1 and X_2 is 9 if $X_1 = 3$, then X_2 is equal to:
(a) 3 (b) 9
(c) 27 (d) 81

90. If the two observations are a = 2 and b = – 2, then their geometric mean will be:
(a) zero (b) infinity
(c) impossible (d) negative

91. Geometric mean of – 4, – 2 and 8 is:
(a) 4 (b) 0
(c) – 2 (d) impossible.

92. The ratio among the number of items and the sum of reciprocals of items, is called:
(a) arithmetic mean (b) geometric mean
(c) harmonic mean (d) mode

93. Harmonic mean for X_1 and X_2 is:
(a) $\frac{2}{X_1 + X_2}$ (b) $\frac{2X_1X_2}{X_1 + X_2}$
(c) $\frac{X_1 + X_2}{2}$ (d) $\frac{X_1 + X_2}{2X_1X_2}$

94. The appropriate average for calculating the average speed of a journey is:
(a) median (b) arithmetic mean
(c) mode (d) harmonic mean

95. Harmonic mean gives less weightage to:
(a) small values (b) large values
(c) positive values (d) negative values

96. The harmonic mean of the values 5, 9, 11, 0, 17, 13 is:
(a) 9.5 (b) 6.2
(c) 0 (d) impossible.

97. If the harmonic mean of the two numbers X_1 and X_2 is 6.4 if $X_2 = 16$, then X_1 is:
(a) 4 (b) 10
(c) 16 (d) 20

98. If a = 5 and b = – 5, then their harmonic mean is:
(a) – 5 (b) 5
(c) 0 (d) ∞

99. For an open-end frequency distribution, it is not possible to find:
(a) arithmetic mean (b) geometric mean
(c) harmonic mean (d) all of the above

100. If all the items in a variable are non zero and non negative then:
(a) A.M > G.M > H.M (b) G.M > A.M > H.M
(c) H.M > G.M > A.M (d) A.M < G.M < H.M

101. The geometric mean of a set of positive numbers $X_1, X_2, X_3, \ldots, X_n$ is less than or equal to their arithmetic mean but is greater than or equal to their:
(a) harmonic mean (b) median
(c) mode (d) lower and upper quartiles

102. Geometric mean and harmonic mean for the values 3, – 11, 0, 63, – 14, 100 are:
(a) 0 and 3 (b) 3 and – 3
(c) 0 and 0 (d) impossible

103. If the arithmetic mean and harmonic mean of two positive numbers are 16 and 4, then their geometric mean will be:
(a) 4 (b) 8
(c) 16 (d) 64

104. The arithmetic mean and geometric mean of two observations are 16 and 8 respectively, then harmonic mean of these two observations is:
(a) 4 (b) 8
(c) 16 (d) 32

105. The geometric mean and harmonic mean of two values are 8 and 4 respectively, then arithmetic mean of values is:
(a) 4 (b) 16
(c) 24 (d) 128

106. Which pair of averages cannot be calculated when one of numbers in the series is zero?
(a) Geometric mean and median (b) Harmonic mean and mode
(c) Simple mean and weighted mean (d) Geometric mean and harmonic mean

107. In a given data the average which has the least value is:
   (a) mean. (b) median.
   (c) harmonic mean. (d) geometric mean

108. If all the values in a series are same, then:
   (a) A.M = G.M = H.M (b) A.M ≠ G.M ≠ H.M
   (c) A.M > G.M > H.M (d) A.M < G.M < H.M

109. The averages are effected by change of:
   (a) origin (b) scale
   (c) both (a) and (b) (d) none of the above

Answers

1.	(c)	2.	(c)	3.	(b)	4.	(a)	5.	(b)	6.	(a)	7.	(b)	8.	(c)
9.	(d)	10.	(d)	11.	(b)	12.	(c)	13.	(b)	14.	(a)	15.	(a)	16.	(b)
17.	(c)	18.	(c)	19.	(b)	20.	(d)	21.	(c)	22.	(a)	23.	(a)	24.	(c)
25.	(d)	26.	(c)	27.	(a)	28.	(d)	29.	(b)	30.	(d)	31.	(c)	32.	(a)
33.	(c)	34.	(d)	35.	(c)	36.	(c)	37.	(c)	38.	(c)	39.	(b)	40.	(d)
41.	(b)	42.	(a)	43.	(b)	44.	(b)	45.	(b)	46.	(c)	47.	(d)	48.	(a)
49.	(b)	50.	(c)	51.	(a)	52.	(c)	53.	(d)	54.	(c)	55.	(d)	56.	(d)
57.	(b)	58.	(b)	59.	(c)	60.	(c)	61.	(c)	62.	(a)	63.	(b)	64.	(d)
65.	(c)	66.	(c)	67.	(a)	68.	(a)	69.	(c)	70.	(b)	71.	(a)	72.	(a)
73.	(c)	74.	(a)	75.	(c)	76.	(d)	77.	(a)	78.	(d)	79.	(a)	80.	(b)
81.	(c)	82.	(a)	83.	(c)	84.	(c)	85.	(c)	86.	(b)	87.	(b)	88.	(d)
89.	(c)	90.	(c)	91.	(d)	92.	(c)	93.	(b)	94.	(d)	95.	(b)	96.	(d)
97.	(a)	98.	(d)	99.	(d)	100	(a)	101	(a)	102	(d)	103	(b)	104	(a)
105	(b)	106	(d)	107	(c)	108	(a)	109	(c)						

SHORT QUESTIONS

Q.1 What is meant by measures of central tendency?

Q.2 What are the characteristics of a good measure of central tendency?

Q.3 What is a statistical average?

Q.4 Why do we calculate averages?

Q.5 What are the desirable properties for an average?

or

Write down the qualities of a good average.

Q.6 Why an average is called measure of central tendency?

Q.7 Define the importance of average in practical life.

Q.8 Name the important types of averages.

or

Write down the different measures of averages.

Q.9 Describe the uses of various averages in practical life.

Q.10 Define arithmetic mean.

Q.11 Define the population mean.

Q.12 Define the sample mean.

Q.13 Write down the various methods of calculating arithmetic mean.

Q.14 Write down the mathematical properties of arithmetic mean.

Q.15 Write down the advantages of arithmetic mean.

Q.16 Write down the demerits of arithmetic mean.

Q.17 Given $\Sigma(X - 10) = 2.8$ and $n = 5$. Find the sample mean.

Ans. 10.56

Q.18 If the sum of deviations from $X = 15$ for 10 values is 25, find the mean.

Ans. 17.5

Q.19 The daily wages of 5 workers are Rs. 200, 300, 400, 500, 600 and the mean of these wages is Rs. 400. Find the mean after increasing the wages by 10%.

Ans. Rs. 440

Q.20 The mean of 10 values is 15. If one more value is included, the mean becomes 18. Find the included value.

Ans. 48

Q.21 The mean of 3 numbers is 20. If one more number is included, the mean becomes 25. Find the included number.

Ans. 40

Q.22 The mean of 5 observations is 60. Another item is included in the observations and now the mean becomes 62. Find the included item.

Ans. 72

Q.23 A student calculated the value of mean as 20 from 25 observations. It was later discovered at the time of checking that he had copied down two values as 7 and 18 while the correct values where 13 and 17. Find the correct value of the mean.

Ans. 20.2

Q.24 The mean of n values is 10. If a new value 20 is excluded, the mean becomes 8. Find the value of n.

Ans. 6

Q.25 Given $u = \dfrac{X - 170}{5}$, $\Sigma fu = 100$ and $\Sigma f = 200$. Find arithmetic mean.

Ans. 172.5

Q.26 The average IQ of 10 students in a Statistics course is 114. If 9 of the students have IQs of 101, 125, 118, 128, 106, 115, 99, 118 and 109, what must be the IQ of 10th student?

Ans. 121

Q.27 A given stock was purchased at the following prices at various times. 20 shares at Rs. 8.20 a share, 100 shares at Rs. 10.90 a share, 50 shares at Rs. 9.40 a share, 200 shares at Rs. 7.80 a share. Find the mean cost per share.

Ans. 8.88

Q.28 Show that the sum of deviations from mean is zero.

Ans. $\sum_{i=1}^{n} (X_i - \bar{X}) = 0.$

Q.29 Verify that $\sum_{i=1}^{n} f_i (X_i - \bar{X}) = 0.$

Ans. $\sum_{i=1}^{n} f_i (X_i - \bar{X}) = 0.$

Q.30 For a certain distribution $\Sigma (X - 21) = 18$ and $\Sigma (X - 28) = -24$.

Find the value of n and $\bar{X}$.

Ans. $n = 6$ and $\bar{X} = 24$.

Q.31 Write short note on weighted arithmetic mean. or Define weighted mean.

Q.32 Given $X = 1, 2, 3, \cdots, n$ and $W = 1, 2, 3, \cdots, n$. Find the weighted mean.

Ans. $(2n + 1)/3$

Q.33. Differentiate between simple arithmetic mean and weighted arithmetic mean.

Q.34 Define combined arithmetic mean.

Q.35 What is the formula of combined arithmetic mean?

Q.36 Given $\bar{X}_1 = 15$, $\bar{X}_2 = 25$, $\bar{X}_3 = 35$ and $\bar{X}_4 = 5$. Each mean is based on the same number of observations, find the overall mean.

Ans. 20

Q.37 Given $\bar{X}_1 = 2.53$, $\bar{X}_2 = 1.97$, $\bar{X}_3 = 4.05$, $\bar{X}_4 = 3.33$ and each mean is based on eight values. Compute $\bar{X}_c$.

Ans. 2.97

Q.38 The arithmetic mean of 100 items is 5 and the arithmetic mean of 150 items is 4.8. Find combined mean.

Ans. 4.88

Q.39 Average weight of 10 students is 60 kg. and average weight of 15 students is 64 kg. Find the average weight of all the 25 students.

Ans. 62.4 kg.

Q.40 The mean wage of 100 workers working in a factory running two shifts of 70 and 30 workers respectively is Rs. 84. The mean wage of 70 workers working in morning shift is Rs. 90. Find the mean wage of 30 workers working in the evening shift.

Ans. Rs. 70

Q.41 Define the term median.

Q.42 Enlist any four advantages of median.

Q.43 Write down any four demerits of median.

Q.44 Given $l = 60$, $h = 10$, $f = 20$, $n = 80$ and $c = 30$. Find median.

Ans. 65

Q.45 Define the quartiles. Also write down its formulas.

Q.46 Define the deciles and explain how to compute.

Q.47 Write a short note on percentiles.

Q.48 Can all quartiles be expressed as percentiles? Explain.

Ans. $Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$

Q.49 Define mode. or Describe the definition of mode.

Q.50 Give important advantages and disadvantages of the mode.

Q.51 Given $l = 200$, $f_m = 25$, $f_1 = 20$, $f_2 = 20$ and $h = 10$. Find mode.

Ans. 205

Q.52 Write down the empirical relationship be' on mean, median and mode.

Q.53 What is the relative position of the mean, median and mode in the symmetrical and skewed distributions?

Q.54 Write the formulas of mean, median and mode in case of a frequency distribution.

Q.55 In a moderately skewed distribution, the value of the mean and median are 120 and 110 respectively. Find the value of the mode.

Ans. 90

Q.56 In a moderately skewed distribution, mean = 25 and mode = 28. Find the value of the median.

Ans. 26

Q.57 In moderately asymmetrical distribution, the value of median is 42 and the value of mode is 40. Find the mean.

Ans. 43

Q.58 Find out the missing figure:

(a) $\text{Mean} = \frac{1}{2}(\ ?\ \text{median} - \text{mode})$

(b) $\text{Median} = \ ?\ (2\ \text{mean} + \text{mode})$

(c) $\text{Mode} = 3\ \text{median} - \ ?\ \text{mean}$

Ans. (a) 3 (b) 1/3 (c) 2

Q.59 Define geometric mean.

Q.60 Write down the properties of geometric mean.

Q.61 Write down any four advantages of geometric mean.

Q.62 Give any four demerits of geometric mean.

Q.63 The geometric mean of a series of 4 items is 10.2, find the product of all the items.

Ans. 10824.3216

Q.64 The geometric mean of two positive numbers is 10. By including the third number, the geometric mean becomes 8. What is the third number?

Ans. 5.12

Q.65 The geometric mean of a series of 3 values is 4. By including the 4th value, the geometric mean becomes 2. What is the 4th value?

Ans. 0.25

Q.66 Write short note on harmonic mean.

Q.67 Write down the advantages and disadvantages of harmonic mean.

Q.68 Explain how to compute geometric mean and harmonic mean.

Q.69 Explain the geometric mean and write down the relationship between arithmetic mean, geometric mean and harmonic mean.

Q.70 Write the formulas of arithmetic mean, geometric mean and harmonic mean of a frequency distribution.

Q.71 Given $X_1 = 4$, $X_2 = 4$ and $X_3 = 4$. Show that A.M. = G.M. = H.M.

Ans. A.M. = G.M. = H.M. = 4

Q.72 If $X_1 = 2$ and $X_2 = 8$, show that A.M. > G.M. > H.M.

Ans. $5 > 4 > 3.2$

Q.73 Show that the geometric mean of any two values is equal to the geometric mean of their arithmetic mean and harmonic mean.

Q.74 The geometric mean and harmonic mean of two values are 20 and 16 respectively. Find arithmetic mean of these values.

Ans. 25

Q.75 The arithmetic mean and harmonic mean of two positive numbers are 25 and 16 respectively. Find their geometric mean.

Ans. 20

Q.76 The arithmetic mean and geometric mean of two observations are 25 and 20 respectively. Find harmonic mean of these two observations.

Ans. 16

Q.77 A student computed the values of arithmetic mean, geometric mean and harmonic mean of a continuous variable as 40,32 and 50. Identify the values of arithmetic mean, geometric mean and harmonic mean.

Ans. A.M. = 50, G.M. = 40, H.M. = 30

Q.78 Given $X_1 = 4$ and $X_2 = 16$. Show that $G.M = \sqrt{A.M \times H.M.}$

Ans. $G.M = \sqrt{A.M \times H.M} = 8$

Q.79 Given $X_1 = 3$ and $X_2 = 27$. Show that $H.M = (G.M.)^2 / A.M$

Ans. $H.M. = (G.M.)^2 / A.M = 5.4$

Q.80 Write down the relative position of averages.

Q.81 Write down the formulas for different averages.

EXERCISES

Q.1 Find the mean temperature of the following temperatures recorded in an experiment in the laboratory : 81, 94, 64, 80, 75, 69, 96, 66, 80, 91, 85, 79.

Ans. 80

2. The following table gives the hourly income of ten operators in a machine tool factory. Find the arithmetic mean.

Name of operator	A	B	C	D	E	F	G	H	I	J
Income (Rs.)	12	15	18	20	25	30	22	35	37	26

Ans. 24

Q.3 (i) Compute the mean of the following sample values:

1.3, 7.0, 3.6, 4.1, 5.0

Show that sum of the deviations of values from their mean is zero.

(ii) A variable Y is determined from a variable X by the equation $Y = 10 - 4X$.

Find Y when $X = -3, -2, -1, 0, 1, 2, 3, 4, 5$ and show that $\bar{Y} = 10 - 4\bar{X}$.

Ans: (i) 4.2 (ii) $\bar{X} = 1$, $\bar{Y} = 6$

Q.4 The reciprocals of 11 values of X are given below. Calculate the arithmetic mean of X : 0.0500, 0.0454, 0.0400, 0.0333, 0.0285, 0.0232, 0.0213, 0.0200, 0.0182, 0.0151, 0.0143.

Ans. 42.091

Q.5 The logarithms of 10 values of X are: 1.8062, 1.2304, 1.6532, 1.5798, 1.4314, 0.7782, 1.6812, 1.0414, 1.7559, 1.5315. Calculate arithmetic mean of X values.

Ans. 34.7

Q.6 A student calculated the value of mean as 20 from 25 observations. It was later discovered at the time of checking that he had copied down two values as 7 and 18 while the correct values were 13 and 17. Find the correct value of the arithmetic mean.

Ans. 20.2

Q.7 The mean of the 10 numbers is 8. If an eleventh number is included in the data, the mean becomes 9. What is the value of the eleventh number?

Ans. 19

Q.8 Arithmetic mean of 98 items is 50. Two items 60 and 70 were left out at the time of calculations. Find the mean of 100 items?

Ans. 50.3

Q.9 From the following frequency distribution find out mean height of the students.

Height (inches)	64	65	66	67	68	69	70	71	72	73
No. of students	4	9	12	18	20	12	10	9	4	2

Ans. 68

Q.10 Show that sum of the deviations of values from their mean is zero.

Classes	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59
Frequency	5	25	40	20	10

Ans. $\bar{X} = 35$, $\Sigma f(X - \bar{X}) = 0$

Q.11 Calculate the average marks from the following data by

(i) the direct method (ii) the short-cut method.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of students	5	12	15	25	8	3	2

Ans. (i) 30.14 (ii) 30.14

Q.12 The following table gives the daily wages of a group of persons working in a factory. Apply the step deviation method to find the mean daily wages.

Daily wages (in Rs.)	No. of persons	Daily wages (in Rs.)	No. of persons
155 – 157	4	170 – 172	62
158 – 160	8	173 – 175	48
161 – 163	26	176 – 178	14
164 – 166	53	179 – 181	6
167 – 169	89		

Ans. 168.76

Q.13 A student obtained 40, 50, 60, 80 and 45 marks in the subjects of English, Urdu, Mathematics, Statistics and Physics respectively. Assigning weights of 5, 2, 4, 3 and 1 respectively for the above mentioned subjects. Find weighted arithmetic mean per subject.

Ans. 55

Q.14 The following table indicates the increase in cost of living over July 1992, for a working class family as at 1st July 1996, and the weights assigned to various groups.

Groups	Percentage increase	Weights
Food	29	7.5
Rent	54	2.5
Clothing	97.5	1.5
Fuel and lighting	75	1.0
Other items	75	0.5

Find out the weighted average of the increase in cost of living.

Ans. 47.02

Q.15 An examination was held to decide the award of a scholarship. The weights of various subjects were different. The marks obtained by 3 candidates (out of 100 in each subject) are given below:

Subject	Weight	Marks A	Marks B	Marks C
Statistics	4	63	60	65
Mathematics	3	65	64	70
Economics	2	58	56	63
Punjabi	1	70	86	52

If the candidate getting the highest marks is to be awarded the scholarship, who should get it?

Ans. Student C.

Q.16 The arithmetic mean of 100 items is 5 and the arithmetic mean of 150 items is 4.8. Find the combined arithmetic mean.

Ans. 4.88

Q.17 The teachers of statistics reported mean examination marks of 37.5, 41 and 42 in their classes which consisted of 32, 25 and 17 students respectively. Determine the mean marks for all the classes taken together.

Ans. 39.72.

Q.18 A distribution consists of four components with frequencies 15, 12, 16, 21, having their means 16.5, 20.3, 21.6 and 26.2. Find the mean of the combined distribution.

Ans. 21.67

Q.19 The mean heights and the number of students in three sections of a statistics class are given below:

Section	A	B	C
No. of boys	30	37	43
Mean height	62"	58"	61"

Find the combined mean of the whole class.

Ans. 60.26

Q.20 The mean height of 25 students in section A of a class is 61" and the mean height of 35 students in section B of the same class is 58". Find the overall mean height of the 60 students.

Ans. 59.25

Q.21 The mean wage of 100 workers working in a factory running two shifts of 70 and 30 workers respectively is Rs. 84. The mean wage of 70 workers working in morning shift is Rs. 90. Find the mean wage of 30 workers working in the evening shift.

Ans. Rs. 70

Q.22 The mean of marks in mathematics of 100 students of a class was 72. The mean of marks of boys was 75 while their number was 70. Find out the mean marks of girls in the class.

Ans. 65

Q.23 The average wage of 4 men is Rs. 17 per hour. What is the average wage of further 6 men if the average wage of all 10 men is Rs. 20?

Ans. 22

Q.24 The mean marks obtained by 300 students are 56. The mean of the top 100 students of them was found to be 80 and the mean of the bottom 100 of them was found to be 22. What is the mean of the remaining 100 students.

Ans. 66

Q.25 From the following data find the missing frequency when mean is 15.38:

Mid value	10	12	14	16	18	20
No. of students	3	7	?	20	8	5

Ans. 12

Q.26 The arithmetic mean of the following series is 30.5 marks. Find the missing figure.

Marks	10	20	?	40	50
Frequency	8	10	20	15	7

Ans. 30

Q.27 Find out the median of the following values:

(i) 8, 10, 12, 18, 20, 27, 32

(ii) 5, 4, 8, 3, 7, 2, 9

(iii) 18.3, 20.6, 19.3, 22.4, 20.2, 18.8, 19.7, 20.0

(iv) 22, 26, 14, 30, 18, 11, 35, 41, 12, 32

Ans. (i) 18 (ii) 5 (iii) 19.85 (iv) 24

Q.28 Find the value of the median.

Hourly wages (Rs.)	40 – 60	60 – 80	80 – 100	100 – 120	120 – 140	140 – 160	160 – 180
No. of employees	13	23	101	182	105	19	7

Ans. 109.67

Q.29 Compute median, 6th decile and 74th percentile for the following data:

Diameter (inches)	Frequency	Diameter (inches)	Frequency
0.7312 – 0.7313	10	0.7320 – 0.7321	30
0.7314 – 0.7315	15	0.7322 – 0.7323	8
0.7316 – 0.7317	20	0.7324 – 0.7325	2
0.7318 – 0.7319	25		

Ans. Median = 0.73183, $D_6 = 0.73192$, $P_{74} = 0.73203$

Q.30 The deviations from X = 22.5 of different values of X are: – 12, – 8.5, 3.0, 0, 2.5, 6.6, 9.2, 1.6, 0.5 and 0.4. Find the lower and the upper quartiles of variable X.

Ans. $Q_1 = 20.375$, $Q_3 = 26.4$

Q.31 Find median, upper and lower quartiles and 65th percentile for the distribution of heights of 300 boys.

Height in inches	59	58	57	56	55	54	53	52	51	50	49	48	47
No. of boys	1	3	7	8	25	30	55	50	40	38	30	9	4

Ans. Median = 52.08, $Q_1 = 50.34$, $Q_3 = 53.48$, $P_{65} = 52.94$

Q.32 The daily wages for a group of 200 persons have been obtained from a frequency distribution of a continuous variable X, after making the substitution u = (X – 130) /20.

$u = \dfrac{X-130}{20}$	– 2	– 1	0	1	2
No. of persons	7	50	80	60	3

Calculate the median, lower and upper quartiles, 3rd decile and 98th percentile.

Ans. Median = 130.75, $Q_1 = 117.2$, $Q_3 = 144.33$, $D_3 = 120.75$, $P_{98} = 159.67$

Q.33 Calculate the median, 3rd decile and 20th percentile from the following data:

Central size	2.5	7.5	12.5	17.5	22.5
Frequency	7	18	25	30	20

Ans. Median = 15, $D_3 = 11$, $P_{20} = 8.61$

Q.34 Find the mode of each of the following sets of values.

(i) 27, 29, 27, 25, 24, 27, 25, 29 (ii) 412, 426, 435, 412, 428, 435, 427

(iii) 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 18, 12 (iv) 101, 106, 99, 108, 76, 87, 102, 93

(v) 9, 12, 5, 4, 3, 6, 11, 7, 5, 2, 11, 9, 13, 7, 6, 8.

Ans. (i) 27 (ii) 412 and 435 (iii) 9 (iv) no mode (v) no mode

Q.35 The following table, shows the number of men in various age groups with some form of paid employment in a village. The age recorded for each man is the number of completed years lived.

Age(years)	11 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 – 70	71 – 80
Frequency	12	14	26	35	23	5	1

Calculate the mode.

Ans. 44.79

Q.36 Compute mode of the data:

Hourly wages	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14	14 – 16
No. of employees	13	111	182	105	19	7

Ans. 8.96

Q.37 Following is the distribution of the size of certain farms selected at random from a district, calculate the mode of distribution:

Central size of the farm in acres	10	20	30	40	50	60	70
Number of farms	7	12	17	29	31	5	3

Ans. 45.71

Q.38 The following table shows the distributions of the maximum loads in short tons supported by certain cables produced by a company:

Maximum load	No. of cables	Maximum load	No. of cables
9.3 – 9.7	2	11.3 – 11.7	14
9.8 – 10.2	5	11.8 – 12.2	6
10.3 – 10.7	12	12.3 – 12.7	3
10.8 – 11.2	17	12.8 – 13.2	1

Determine mean, median and mode.

Ans. Mean = 11.092, Median = 11.073, Mode = 11.063

Q.39 Find the value of mode by using the empirical relationship between averages for the following data:

Marks	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59
No. of students	5	25	40	20	10

Ans. Mode = 33.5

Q.40 (i) In a moderately asymmetrical distribution the mode and mean are 32.1 and 35.4 respectively. Calculate the median.

(ii) In a moderately asymmetrical series, the value of arithmetic mean and median is 20 and 13.67 respectively. Find out the value of mode.

(iii) If the mode and mean of a moderately asymmetrical series are respectively 16 cm. and 20.2 cm., compute the most probable value of median.

Ans. (i) 34.3 (ii) 16.01 (iii) 18.8 cm.

Q.41 Form a frequency distribution taking the variable as the number of letters in a word in the stanza given below and calculate the mean, the median and mode of the distribution.

"Heights by great men reached and kept,
Were not attained by sudden flight;
But they while their companions slept,
Were boiling upwards in the night".

Ans. Mean = 4.8, Median = 5, Mode is illdefined

Q.42 The expenditure of 100 families is given below :

Expenditure	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of families	14	?	27	?	15

Mode for the distribution is 24. Calculate the missing frequencies.

Ans. 23 and 21

Q.43 Calculate the geometric mean of the following price relatives:

Commodity	Wheat	Rice	Pulses	Sugar	Salt	Oils
Price relatives	207	198	156	124	107	196

Ans. 159.77

Q.44 Calculate geometric mean for the following numbers:

(i) 8, 40, 175, 1209, 2000

(ii) 6.5, 169.0, 11.0, 112.5, 14.2, 75.0, 35.5, 215.0

(iii) 0.35, 3 , 30, 300, 3000, 0.3, 0.03, 0.0362, 0.00482, 58642

Ans. (i) 168.39 (ii) 42.70 (iii) 4.38

Q.45 The reciprocals of 8 different values of X are 0.0667, 0.1111, 0.0833, 0.0556, 0.0500, 0.0357, 0.0278, 0.0222. Calculate (i) arithmetic mean (ii) geometric mean of X values.

Ans. (i) 22.875 (ii) 20.082

Q.46 The deviations of 10 different values from X = 22 are: 0, 2, – 3, – 4, 6, 5, 8, – 1, 0, 3. Calculate the geometric mean.

Ans. 23.31

Q.47 For the following data find the geometric mean.

Weight in gm.	60–80	80–100	100–120	120–140	140–160	160–180	180–200
No. of apples	5	14	17	10	1	0	2

Ans. 105.61

Q.48 Find geometric mean from the following frequency distribution:

X	2	3	4	5	6
f	5	7	8	3	2

Ans. 3.41

Q.49 Find two numbers whose mean is 9.0 and geometric mean is 7.2.

Ans. 14.4 and 3.6

Q.50 The mean and geometric mean of three numbers are 7 and 4 respectively. Find all the three numbers if mean of first two numbers is 10.

Ans. 16, 4 and 1

Q.51 The mean and geometric mean of three numbers are 34 and 18 respectively. Find all the three numbers when the geometric mean of first two numbers is 9.

Ans. 27, 3 and 72

Q.52 A man gets a rise of 15 % in his salary after the first year of his service and a further rise of 20 % and 25 % in his salary at the end of his second and third year of service respectively. Find his average annual percentage increase.

Ans. 19.92 %.

Q.53 All children in a number of families were together at a party. Each child was asked to state the number of children in his (or her) family. Two children said one child, four children said two children and twelve children said three.

(i) How many families were at the party?

(ii) Calculate the arithmetic mean of the number of children per family.

Ans. (i) 8 (ii) 2.25 (2 children approximately)

Q.54 The population of a country increased by 20 % in the first decade, 30 % in the second decade and 45 % in the third decade. What is the average rate of increase per decade in the population?

Ans. 31.28 %

Q.55 If the industrial production increased 3 percent in first year, 4 percent in second year and 5 percent in third year, what is the average annual increase for the three years.

Ans. 4 %

Q.56 The arithmetic mean of two items is 12.5 and geometric mean is 10. Find two items.

Ans. 20 and 5

Q.57 The weekly incomes of 8 families in rupees in a certain locality are given below. Calculate the harmonic mean.

Family	A	B	C	D	E	F	G	H
Income(Rs.)	2700	2000	5000	7500	1800	2500	4800	4200

Ans. 3071.02

Q.58 Calculate the harmonic mean from the following data :

X	2.1	2.4	2.7	3.0	3.3	3.6	3.9	4.2	4.5
f	27	42	59	69	54	43	21	16	9

Ans. 2.96

Q.59 Calculate harmonic mean of the variable X from the following data :

$u = \dfrac{X-3.5}{0.5}$	-3	-2	-1	0	1	2	3
Frequency	15	38	65	92	80	40	20

Ans. 3.38

Q.60 If the arithmetic mean of two numbers is 20 and their geometric mean is 16, find the harmonic mean.

Ans. 12.8

Q.61 Consider the following data:

Hourly wages (Rs.)	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
Number of persons	4	8	16	8	4

Compute the harmonic mean.

Ans: 63.05

Chapter 4

MEASURES OF DISPERSION

4.1. INTRODUCTION

A modern student of statistics is mainly interested in the study of variability and uncertainty. In this chapter we shall discuss variability and its measures and uncertainty will be discussed in chapter seven. We live in a changing world. Changes are taking place in every sphere of life. A man of statistics does not show much of interest in those things which are constant. The total area of the earth may not be very important to a research minded person but the area under different crops, area covered by forests, area covered by residential and commercial buildings are figures of great importance because these figures keep on changing from time to time and from place to place. Very large number of experts are engaged in the study of changing phenomenon. Experts working in different countries of the world keep a watch on forces which are responsible for bringing changes in the fields of human interest. The agricultural, industrial and mineral production and their transportation from one part to the other parts of the world are the matters of great interest to the economists, statisticians and other experts. The changes in human population, the changes in standard of living, changes in literacy rate and the changes in prices attract the experts to make detailed studies about them and then correlate these changes with the human life. Thus variability or variation is something connected with human life and the study is very important for mankind.

4.2. DISPERSION

The word dispersion has a technical meaning in statistics. The average measures the centre of the data. It is one aspect of the observations. Another feature of the observations is as to how the observations are spread about the centre. The observations may be close to the centre or they may be spread away from the centre. If the observations are close to the centre (usually the arithmetic mean or median), we say that dispersion, scatter or variation is small. If the observations are spread away from the centre, we say dispersion is large. Suppose we have three groups of students who have obtained the following marks in a test. The arithmetic means of the three groups are also given below:

Group A: 46, 48, 50, 52, 54, $\bar{X}_A = 50$

Group B: 30, 40, 50, 60, 70, $\bar{X}_B = 50$

Group C: 40, 50, 60, 70, 80, $\bar{X}_C = 60$

In group A and B arithmetic means are equal i.e., $\bar{X}_A = \bar{X}_B = 50$. But in group A the observations are concentrated on the centre. All the students of group A have almost the same level of performance. We say that there is consistence in the observations in group A. In group B, the mean is 50 but the observations are not close to the centre. One observation is as small as 30 and one observation is as large as 70. Thus there is greater dispersion in group B. In group C the mean is 60 but the spread of the observations with respect to the centre 60 is the same as the spread of the observations in group B with respect to their own centre which is 50. Thus in group B and C the means are different but their dispersion is the same. In group A and C the means are different and their dispersions are also different. Dispersion is an important feature of the observations and it is measured with the help of the measures of dispersion, scatter or variation. The word variability is also used for this idea of dispersion.

The study of dispersion is very important in statistical data. If in a certain factory there is consistence in the wages of workers, the workers will be satisfied. But if some workers have high wages and some have low wages, there will be unrest among the low paid workers and they might go on strikes and arrange demonstrations. If in a certain country some people are very poor and some are very rich, we say there is economic disparity. It means that dispersion is large. The idea of dispersion is important in the study of wages of workers, prices of commodities, standard of living of different people, distribution of wealth, distribution of land among farmers and various other fields of life. Some brief definitions of dispersion are:

(1) The degree to which numerical data tend to spread about an average value is called the dispersion or variation of the data.

(2) Dispersion or variation may be defined as a statistics signifying the extent of the scatter of items around a measure of central tendency.

(3) Dispersion or variation is the measurement of the scatter of the size of the items of a series about the average.

4.3. MEASURES OF DISPERSION

For the study of dispersion, we need some measures which show whether the dispersion is small or large. There are two types of measures of dispersion which are:

(a) Absolute measures of dispersion. (b) Relative measures of dispersion.

4.3.1. ABSOLUTE MEASURES OF DISPERSION

These measures give us an idea about the amount of dispersion in a set of observations. They give the answers in the same units as the units of the original observations. When the observations are in kilograms, the absolute measure is also in kilograms. If we have two sets of observations, we cannot always use the absolute measures to compare their dispersion. We shall explain later as to when the absolute measures can be used for comparison of dispersion in two or more than two sets of data. The absolute measures which are commonly used are:

(i) The range (ii) The quartile deviation

(iii) The mean deviation (iv) The standard deviation or variance.

4.3.2. RELATIVE MEASURES OF DISPERSION

These measures are calculated for the comparison of dispersion in two or more than two sets of observations. These measures are free of the units in which the original data is measured. If the original data is in rupees or kilometers, we do not use these units with relative measures of dispersion. These measures are a sort of ratio and are called coefficients. Each absolute measure of dispersion can be converted into its relative measure. Thus the relative measures of dispersion are :

(i) Coefficient of Range or Range Coefficient of Dispersion.

(ii) Coefficient of Quartile Deviation or Quartile Coefficient of Dispersion.

(iii) Coefficient of Mean Deviation or Mean Coefficient of Dispersion.

(iv) Coefficient of Standard Deviation or Standard Coefficient of Dispersion.

(v) Coefficient of Variation (a special case of Standard Coefficient of Dispersion).

4.4. THE RANGE

Range is defined as the difference between the maximum and the minimum observation of the given data. If X_m denotes the maximum observation, X_0 denotes the minimum observation then the range is defined as

$$\text{Range} = X_m - X_0$$

In case of grouped data, the range is the difference between the upper boundary of the highest class and the lower boundary of the lowest class. It is also calculated by using the difference between the mid points of the highest class and the lowest class. It is the simplest measure of dispersion. It gives a general idea about the total spread of the observations. It does not enjoy any prominent place in statistical theory. But it has its application and utility in quality control methods which are used to maintain the quality of the products produced in factories. The quality of products is to be kept within certain range of values.

The range is based on the two extreme observations. It gives no weight to the central values of the data. It is a poor measure of dispersion and does not give a good picture of the overall spread of the observations with respect to the centre of the observations. Let us consider three groups of data which have the same range :

Group A: 30, 40, 40, 40, 40, 40, 50

Group B: 30, 30, 30, 40, 50, 50, 50

Group C: 30, 35, 40, 40, 40, 45, 50

In all the three groups the range is $50 - 30 = 20$. In group A there is concentration of observations in the centre. In group B the observations are friendly with the extreme corners and in group C the observations are almost equally distributed in the interval from 30 to 50. The range fails to explain these differences in the three groups of data. This defect in range cannot be removed even if we calculate the coefficient of range which is a relative measure of dispersion. If we calculate the range of a sample, we cannot draw any inference about the range of the population.

4.4.1. COEFFICIENT OF RANGE

It is a relative measure of dispersion and is based on the value of range. It is also called range coefficient of dispersion. It is defined as:

$$\text{Coefficient of Range} = \frac{X_m - X_0}{X_m + X_0}$$

The range $X_m - X_0$ is standardised by the total $X_m + X_0$.

Let us take two sets of observations. Set A contains marks of five students in Mathematics out of 25 marks and group B contains marks of the same students in English out of 100 marks

Set A: 10, 15, 18, 20, 20

Set B: 30, 35, 40, 45, 50

The values of range and coefficient of range are calculated as:

	Range	Coefficient of Range
Set A (Mathematics)	$20 - 10 = 10$	$\frac{20-10}{20+10} = 0.33$
Set B (English)	$50 - 30 = 20$	$\frac{50-30}{50+30} = 0.25$

In set A the range is 10 and in set B the range is 20. Apparently it seems as if there is greater dispersion in set B. But this is not true. The range of 20 in set B is for large observations and the range of 10 in set A is for small observations. Thus 20 and 10 cannot be compared directly. Their base is not the same. Marks in Mathematics are out of 25 and marks in English are out of 100. Thus, it makes no sense to compare 10 with 20. When we convert these two values into coefficient of range, we see that coefficient of range for set A is greater than that of set B. Thus there is greater dispersion or variation in set A. The marks of students in English are more stable than their marks in Mathematics.

Example 4.1.

Following are the wages of 8 workers of a factory. Find the range and the coefficient of range.

Wages (in Rs.): 1400, 1450, 1520, 1380, 1485, 1495, 1575, 1440.

Solution:

Here, Largest value $= X_m = 1575$ and Smallest value $= X_0 = 1380$

$$\text{Range} = X_m - X_0 = 1575 - 1380 = 195$$

$$\text{Coefficient of Range} = \frac{X_m - X_0}{X_m + X_0} = \frac{1575 - 1380}{1575 + 1380} = \frac{195}{2955} = 0.066$$

Example 4.2.

Find the range of the weights of the students of a university.

Weights (kilograms)	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
Number of students	5	18	42	27	8

Solution:

Weights (kilograms)	Class boundaries	Mid value	Number of students
60 – 62	59.5 – 62.5	61	5
63 – 65	62.5 – 65.5	64	18
66 – 68	65.5 – 68.5	67	42
69 – 71	68.5 – 71.5	70	27
72 – 74	71.5 – 74.5	73	8

Method 1:

X_m = Upper class boundary of the highest class = 74.5

X_0 = Lower class boundary of the lowest class = 59.5

Range = $X_m - X_0$ = 74.5 – 59.5 = 15 kilograms.

Method 2:

X_m = Mid value of the highest class = 73

X_0 = Mid value of the lowest class = 61

Range = $X_m - X_0$ = 73 – 61 = 12 kilograms.

4.5. QUARTILE DEVIATION

It is based on the lower quartile Q_1 and the upper quartile Q_3. The difference $Q_3 - Q_1$ is called the inter quartile range. The difference $Q_3 - Q_1$ divided by 2 is called semi-inter-quartile range or the quartile deviation. Thus

$$\text{Quartile Deviation (Q.D.)} = \frac{Q_3 - Q_1}{2}$$

The quartile deviation is a slightly better measure of absolute dispersion than the range. But it ignores the observations on the tails. If we take different samples from a population and calculate their quartile deviations, their values are quite likely to be sufficiently different. This is called sampling fluctuation. It is not a popular measure of dispersion. The quartile deviation calculated from the sample data does not help us to draw any conclusion (inference) about the quartile deviation in the population.

4.5.1. COEFFICIENT OF QUARTILE DEVIATION

A relative measure of dispersion based on the *quartile deviation* is called the *coefficient of quartile deviation.* It is defined as

$$\text{Coefficient of Quartile Deviation} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

It is a pure number free of any units of measurement. It can be used for comparing the dispersion in two or more than two sets of data.

Example 4.3

The wheat production (in kgs.) per acre is given as:

1120, 1240, 1320, 1040, 1080, 1200, 1440, 1360, 1680, 1730.

Find the quartile deviation and coefficient of quartile deviation.

Solution:

After arranging the observations in ascending order, we get

1040, 1080, 1120, 1200, 1240, 1320, 1360, 1440, 1680, 1730

$$Q_1 = \text{Value of}\left(\frac{n+1}{4}\right)\text{th item} = \text{Value of}\left(\frac{10+1}{4}\right)\text{th item}$$

$$= \text{Value of (2.75) th item}$$

$$= \text{2nd item} + 0.75\text{ (3rd item} - \text{2nd item)}$$

$$= 1080 + 0.75(1120 - 1080) = 1080 + 30 = 1110$$

$$Q_3 = \text{Value of } \frac{3(n+1)}{4}\text{ th item} = \text{Value of } \frac{3(10+1)}{4}\text{ th item}$$

$$= \text{Value of (8.25) th item}$$

$$= \text{8th item} + 0.25\text{ (9th item} - \text{8th item)}$$

$$= 1440 + 0.25(1680 - 1440) = 1440 + 60 = 1500$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{1500 - 1110}{2} = \frac{390}{2} = 195$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1500 - 1110}{1500 + 1110} = \frac{390}{2610} = 0.15$$

Example 4.4.

Calculate the quartile deviation and coefficient of quartile deviation from the data given below.

Maximum load (short-tons)	No. of cables	Maximum load (short-tons)	No. of cables
9.25 – 9.75	2	11.25 – 11.75	14
9.75 – 10.25	5	11.75 – 12.25	6
10.25 – 10.75	12	12.25 – 12.75	3
10.75 – 11.25	17	12.75 – 13.25	1

Solution:

The necessary calculations are given below:

Maximum Load (Short-tons)	No. of cables (f)	Cumulative frequencies (c.f.)
9.25 – 9.75	2	2
9.75 – 10.25	5	2 + 5 = 7
10.25 – 10.75	12	7 + 12 = 19
10.75 – 11.25	17	19 + 17 = 36
11.25 – 11.75	14	36 + 14 = 50
11.75 – 12.25	6	50 + 6 = 56
12.25 – 12.75	3	56 + 3 = 59
12.75 – 13.25	1	59 + 1 = 60

Q_1 = Value of $\left(\frac{n}{4}\right)$ th item= Value of $\left(\frac{60}{4}\right)$ th item = 15th item

Q_1 lies in the class 10.25 – 10.75.

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

where $l = 10.25$, $h = 0.5$, $f = 12$, $\frac{n}{4} = 15$ and $c = 7$

Thus, $Q_1 = 10.25 + \frac{0.5}{12}(15 - 7) = 10.25 + 0.33 = 10.58$

Q_3 = Value of $\left(\frac{3n}{4}\right)$ th item = Value of $\frac{3(60)}{4}$ th item = 45th item

Q_3 lies in the class 11.25 – 11.75.

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

where $l = 11.25$, $h = 0.5$, $f = 14$, $\frac{3n}{4} = 45$ and $c = 36$

Thus, $Q_3 = 11.25 + \frac{0.5}{14}(45 - 36) = 11.25 + 0.32 = 11.57$

Quartile Deviation (Q.D.) $= \frac{Q_3 - Q_1}{2} = \frac{11.57 - 10.58}{2} = \frac{0.99}{2} = 0.495$

Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{11.57 - 10.58}{11.57 + 10.58} = \frac{0.99}{22.15} = 0.045$

4.6. THE MEAN DEVIATION

The *mean deviation* or the *average deviation* is defined as the mean of the absolute deviations of observations from some suitable average which may be the arithmetic mean, the median or the mode. The difference (X – average) is called deviation and when we ignore the negative sign, this deviation is written as $|X - \text{average}|$ and is read as modulus deviation. The mean of these modulus or absolute deviations is called the mean deviation or the mean absolute deviation.

Thus for a sample data in which the suitable average is the $\bar{X}$, the mean deviation (M.D.) is given by the relation:

$$\text{M.D.} = \frac{\Sigma|X-\bar{X}|}{n}$$

For a frequency distribution, the mean deviation is given by

$$\text{M.D.} = \frac{\Sigma f|X-\bar{X}|}{\Sigma f}$$

when the mean deviation is calculated about the median, the formula becomes

$$\text{M.D. (about median)} = \frac{\Sigma f|X-\text{Median}|}{\Sigma f}$$

The mean deviation about the mode is

$$\text{M.D. (about mode)} = \frac{\Sigma|X-\text{Mode}|}{\Sigma f}$$

For a population data the mean deviation about the population mean μ is

$$\text{M.D.} = \frac{\Sigma f|X-\mu|}{\Sigma f}$$

The mean deviation is a better measure of absolute dispersion than the range and the quartile deviation.

A drawback in the mean deviation is that we use the absolute deviations $|X - \text{average}|$ which does not seem logical. The reason for this is that $\Sigma(X - \bar{X})$ is always equal to zero. Even if we use median or mode in place of $\bar{X}$, even then $\Sigma(X - \text{median})$ or $\Sigma(X - \text{mode})$ may be zero or approximately zero with the result that the mean deviation would always be either zero or close to zero. Thus the very definition of the mean deviation is possible only on the absolute deviations.

The *mean deviation* is based on all the observations, a property which is not possessed by the *range* and the quartile deviation. The formula of the mean deviation gives a mathematical impression that it is a better way of measuring the variation in the data. Any suitable average among the mean, median or mode can be used in its calculation but the value of the mean deviation is minimum if the deviations are taken from the median. A serious drawback of the mean deviation is that it cannot be used in statistical infer

4.6.1. COEFFICIENT OF THE MEAN DEVIATION

A relative measure of dispersion based on the mean deviation is called the *coefficient of the mean deviation* or the mean coefficient of dispersion. It is defined as the ratio of the mean deviation to the average used in the calculation of the mean deviation. Thus

$$\text{Coefficient of M.D. (about mean)} = \frac{\text{Mean Deviation from Mean}}{\text{Mean}}$$

$$\text{Coefficient of M.D.(about median)} = \frac{\text{Mean Deviation from Median}}{\text{Median}}$$

$$\text{Coefficient of M.D.(about mode)} = \frac{\text{Mean Deviation from Mode}}{\text{Mode}}$$

Example 4.5.

Calculate the mean deviation from (i) arithmetic mean (ii) median (iii) mode in respect of the marks obtained by nine students given below and show that the mean deviation from median is minimum.

Marks (out of 25): 7, 4, 10, 9, 15, 12, 7, 9, 7.

Solution:

After arranging the observations in ascending order, we get

Marks: 4, 7, 7, 7, 9, 9, 10, 12, 15.

$$\text{Mean} = \frac{\Sigma X}{n} = \frac{80}{9} = 8.89$$

$$\text{Median} = \text{Value of}\left(\frac{n+1}{2}\right)\text{th item} = \text{Value of}\left(\frac{9+1}{2}\right)\text{th item}$$

$$= \text{Value of (5)th item} = 9$$

Mode = 7(since 7 is repeated maximum number of times)

Marks (X)	|X − Mean|	|X − Median|	|X − Mode|
4	4.89	5	3
7	1.89	2	0
7	1.89	2	0
7	1.89	2	0
9	0.11	0	2
9	0.11	0	2
10	1.11	1	3
12	3.11	3	5
15	6.11	6	8
Total	21.11	21	23

$$\text{M.D. from mean} = \frac{\Sigma \mid X - \text{Mean} \mid}{n} = \frac{21.11}{9} = 2.35$$

$$\text{M.D. from median} = \frac{\Sigma \mid X - \text{Median} \mid}{n} = \frac{21}{9} = 2.33$$

$$\text{M.D. from mode} = \frac{\Sigma \mid X - \text{Mode} \mid}{n} = \frac{23}{9} = 2.56$$

From the above calculations, it is clear that the mean deviation from the median has the least value.

Example 4.6

Compute the mean deviation from median and its coefficient from the following data:

Daily wages (Rs.)	200 – 250	250 – 300	300 – 350	350 – 400	400 – 450
No. of persons	10	20	40	20	10

Solution:

The necessary calculations are given below:

Daily wages (Rs.)	No. of persons f	X	c.f.	|X–median|	f|X–median|
200 – 250	10	225	10	100	1000
250 – 300	20	275	30	50	1000
300 – 350	40	325	70	0	0
350 – 400	20	375	90	50	1000
400 – 450	10	425	100	100	1000
Total	100				4000

$$\text{Median} = \text{Value of} \left(\frac{n}{2}\right) \text{th item} = \text{Value of} \left(\frac{100}{2}\right) \text{th item} = \text{50th item}$$

Median lies in the class 300 – 350

$$\text{Median } l + \frac{h}{f}\left(\frac{n}{2} - c\right) = 300 + \frac{50}{40}(50 - 30) = 300 + 25 = 325$$

$$\text{M.D. from median} = \frac{\Sigma f \mid X - \text{median} \mid}{\Sigma f} = \frac{4000}{100} = 40$$

$$\text{Coefficient of M.D. from median} = \frac{\text{M.D. from median}}{\text{Median}} = \frac{40}{325} = 0.123$$

4.7. THE VARIANCE

The variance of the observations is defined as the mean of the squares of deviations of observations taken from the mean of the observations. For a sample data the variance is denoted by S^2 and the population variance is denoted by σ^2 (sigma square). By definition, the sample variance S^2 has the formula

$S^2 = \frac{1}{n}\Sigma(X - \bar{X})^2$, where $\bar{X}$ is the sample mean and n is the number of observations in the sample. The population variance σ^2 is defined as $\sigma^2 = \frac{1}{N}\ \Sigma(X - \mu)^2$ where μ is the mean of the population and N is the number of observations in the data. It may be remembered that the population variance σ^2 is usually not calculated. The sample variance S^2 is calculated and if needed, this S^2 is used to make inference about the population variance.

The term $\Sigma(X - \bar{X})^2$ is positive, therefore S^2 is always positive. If the original observations are in cms, the value of the variance will be $(\text{cms})^2$. Thus the unit of S^2 is the square of the units of the original measurement. For a frequency distribution the sample variance S^2 and population variance σ^2 are defined as:

$$S^2 = \frac{1}{\Sigma f}\ \Sigma f(X - \bar{X})^2 \qquad \text{and} \qquad \sigma^2 = \frac{1}{\Sigma f}\ \Sigma f(X - \mu)^2$$

4.8. STANDARD DEVIATION

The standard deviation is defined as the positive square root of the mean of the squares of the deviations of observations from their mean. For the sample data the standard deviation is denoted by S and is defined as:

$$S = \sqrt{\frac{1}{n}\Sigma(X - \bar{X})^2}$$

For a population data the standard deviation is denoted by σ(sigma) and is defined as:

$$\sigma = \sqrt{\frac{1}{N}\Sigma(X - \mu)^2}$$

For a frequency distribution the formulas become

$$S = \sqrt{\frac{1}{\Sigma f}\ \Sigma f(X - \bar{X})^2} \qquad \text{and} \qquad \sigma = \sqrt{\frac{1}{\Sigma f}\Sigma f(X - \mu)^2}$$

Thus the standard deviation is the positive square root of the variance. The standard deviation is in the same units as the units of the original observations. If the original observations are in gms, the value of the standard deviation will also be in gms.

The standard deviation plays a dominating role for the study of variation in the data. It is a very widely used measure of dispersion. It stands like a tower among other measures of dispersion. As far as the important statistical tools are concerned, the first important tool is the mean $\bar{X}$ and the second important tool is the standard deviation S. It is based on all the observations and is subject to mathematical treatment. It is of great importance for the analysis of data and for the various statistical inferences.

4.8.1. CHANGE OF ORIGIN

The standard deviation has a strange property that its value does not change if some constant is added to all the observations or some constant is subtracted from all the observations. We can say that the standard deviation and the variance are independent of change of origin. The standard deviation of 1, 2, 3, 4, 5 will be equal to the standard deviation of 101, 102, 103, 104 and 105. The reason is that the difference of 1, 2, 3, 4, 5 from their mean which is 3 is the same as the difference of 101, 102, 103, 104 and 105 from their own mean which is 103.

If $Y = X + a$ where 'a' is an arbitrary constant then $\bar{Y} = \bar{X} + a$ and

$$S.D.(Y) = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n}} = \sqrt{\frac{\Sigma[X+a-\bar{X}-a]^2}{n}} = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}} = S.D.(X) = S_x$$

Thus $S.D.(X+a) = S.D.(X)$ and $Var(X+a) = Var(X)$

Similarly, $S.D.(X-a) = S.D.(X)$ and $Var(X-a) = Var(X)$

4.8.2. CHANGE OF SCALE

When all the observations are multiplied by a certain constant or divided by some constant, the process is called change of scale. Let us multiply X by a constant 'b'.

If $Y = bX$ then $\bar{Y} = b\bar{X}$ and $S.D.(Y) = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n}}$

Putting the values of Y and $\bar{Y}$, we get

$$S.D.(Y) = \sqrt{\frac{\Sigma[bX-b\bar{X}]^2}{n}} = \sqrt{\frac{\Sigma(b^2X^2+b^2\bar{X}^2-2b^2X\bar{X})}{n}}$$

$$= \sqrt{\frac{b^2\,\Sigma(X^2+\bar{X}^2-2X\bar{X})}{n}} = b\sqrt{\frac{\Sigma(X-\bar{X})^2}{n}} = b\,S.D.(X)$$

Therefore $S.D.(Y) = S.D.(bX) = b\,S.D.(X)$ and $Var(bX) = b^2\,Var(X)$

4.8.3. CHANGE OF ORIGIN AND SCALE

Both the operations of change of origin and scale can be applied at the same time. Thus if we take $X-a$, it is change of origin and if we take $\frac{X-a}{h}$, it is called change of origin and scale.

Let $\frac{X-a}{h}$ be denoted by u where u is called the coded variable.

Thus $u = \frac{X-a}{h}$ and $X = a + hu$ and $\bar{X} = a + h\bar{u}$

By definition

$$S.D.(X) = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}.$$ Putting the values of X and $\bar{X}$, we get

$$S.D.(X) = \sqrt{\frac{\Sigma(a+hu-a-h\bar{u})^2}{n}} = h\sqrt{\frac{\Sigma(u-\bar{u})^2}{n}} = h\,S.D.(u)$$

Similarly $\text{Var}(X) = h^2\,\text{Var}(u)$

4.8.4. VARIANCE OF A CONSTANT

Let all the n observations be equal to a constant, say C. Then $\bar{C} = C$

By definition $S.D.(C) = \sqrt{\frac{\Sigma(C-\bar{C})^2}{n}} = \sqrt{\frac{\Sigma(C-C)^2}{n}} = 0$. Thus $\text{Var}(C) = 0$

4.8.5. OTHER FORMULAS FOR CALCULATION OF S AND S^2

By definition $S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$ or $S = \sqrt{\frac{\Sigma f(X-\bar{X})^2}{\Sigma f}}$ (for frequency distribution)

These formulas should be used when $\bar{X}$ has some exact value like $\bar{X} = \frac{20}{5} = 4$ or $\bar{X} = \frac{22}{5} = 4.5$. If $\bar{X} = \frac{20}{6} = 3.3333$, which is non-terminating, there will be some error and also some inconvenience in the calculation of S. The formula for S can be written in some other forms in which we do not have to take the deviations from $\bar{X}$.

Some other formulas for the calculation of S are obtained as :

(i)
$$S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}} = \sqrt{\frac{\Sigma[X^2+\bar{X}^2-2X\bar{X}]}{n}}$$

$$= \sqrt{\frac{\Sigma X^2 + n\bar{X}^2 - 2\bar{X}\,\Sigma X}{n}} = \sqrt{\frac{\Sigma X^2}{n} + \bar{X}^2 - 2\bar{X}\,.\,\bar{X}}$$

$$= \sqrt{\frac{\Sigma X^2}{n} - \bar{X}^2} = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2}$$

Thus $$S = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} \quad \text{and} \quad S^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 = \frac{\Sigma X^2}{n} - \bar{X}^2$$

Thus the variance is the mean of the squares of observations minus the square of the mean. The above expression for ΣX^2 can be written as

$$\Sigma X^2 = n(S^2 + \bar{X}^2)$$

For frequency distribution

$$S = \sqrt{\frac{\Sigma f X^2}{\Sigma f} - \left(\frac{\Sigma f X}{\Sigma f}\right)^2} \quad \text{and} \quad S^2 = \frac{\Sigma f X^2}{\Sigma f} - \left(\frac{\Sigma f X}{\Sigma f}\right)^2$$

It is very convenient to use this formula when the X values are small.

(ii) If we take $D = X - a$, then $\bar{D} = \bar{X} - a$

We have $X = D + a$ and $\bar{X} = a + \bar{D}$

By definition

$$S.D.(X) = S_x = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$$

Putting the values of X and $\bar{X}$, we get

$$S_x = \sqrt{\frac{\Sigma[D+a-a-\bar{D}]^2}{n}} = \sqrt{\frac{\Sigma(D-\bar{D})^2}{n}} = S_D$$

Thus, $S_x = S_D = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2}$ and $S_x^2 = S_D^2 = \frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2$

For frequency distribution the formulas would be

$$S_x = S_D = \sqrt{\frac{\Sigma fD^2}{\Sigma f} - \left(\frac{\Sigma fD}{\Sigma f}\right)^2} \quad \text{and} \quad S_x^2 = S_D^2 = \frac{\Sigma fD^2}{\Sigma f} - \left(\frac{\Sigma fD}{\Sigma f}\right)^2$$

These formulas are based on the property of standard deviation that the standard deviation does not change by change of origin. Standard deviation of X is equal to the standard deviation of D where $D = X \pm a$.

(iii) If we take $u = \frac{X-a}{h}$, then $X = a + hu$ and $\bar{X} = a + h\bar{u}$

By definition $S_x = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$. Putting the values of X and $\bar{X}$ in the formula, we get

$$S_x = \sqrt{\frac{\Sigma(a+hu-a-h\bar{u})^2}{n}} = h\sqrt{\frac{\Sigma(u-\bar{u})^2}{n}}$$

$$S_x = h\sqrt{\frac{\Sigma u^2}{n} - \left(\frac{\Sigma u}{n}\right)^2} \quad \text{and} \quad S_x^2 = h^2\left[\frac{\Sigma u^2}{n} - \left(\frac{\Sigma u}{n}\right)^2\right]$$

For frequency distribution, the formulas for S and S^2 would be

$$S = h\sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2} \quad \text{and} \quad S^2 = h^2\left[\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2\right]$$

These formulas involving the coded variable 'u' are particularly suitable for grouped data. We convert the variable X into another coded variable u where $u = \frac{X-a}{h}$, 'a' is arbitrary constant preferably the value of X somewhere in the middle of the X values. The divisor h is the interval of the grouped data. If the interval is not constant, we can take any suitable value of h.

4.9. COEFFICIENT OF STANDARD DEVIATION

The standard deviation is the absolute measure of dispersion. Its relative measure is called standard coefficient of dispersion or coefficient of standard deviation. It is defined as:

$$\text{Coefficient of Standard Deviation} = \frac{S}{\bar{X}}$$

4.10. COEFFICIENT OF VARIATION

The most important of all the relative measures of dispersion is the coefficient of variation. This word is variation not variance. There is no such thing as coefficient of variance. The coefficient of variation (C.V.) is defined as

$$\text{Coefficient of Variation(C.V.)} = \frac{S}{\bar{X}} \times 100$$

Thus C.V. is the value of S when $\bar{X}$ is assumed equal to 100. It is a pure number and the unit of observations is not mentioned with its value. It is written in percentage form like 20 % or 25 %. When its value is 20 %, it means that when the mean of the observations is assumed equal to 100, their standard deviation will be 20. The C.V. is used to compare the dispersion in different sets of data particularly the data which differ in their means or differ in the units of measurement. The wages of workers may be in rupees and the consumption of meat in their families may be in kgs. The standard deviation of wages in rupees cannot be compared with the standard deviation of amounts of meat in kgs. Both the standard deviations need to be converted into coefficient of variation for comparison. Suppose the value of C.V. for wages is 10 % and the value of C.V. for kgs of meat is 25 %. This means that the wages of workers are consistent. Their wages are close to the overall average of their wages. But the families consume meat in quite different quantities. Some families use very small quantities of meat and some others use large quantities of meat. We say that there is greater variation in their consumption of meat. The observations about the quantity of meat are more dispersed or more variant.

4.11. PROPERTIES OF STANDARD DEVIATION AND VARIANCE

(i) The standard deviation and the variance are positive quantities. The standard deviation is expressed in the same units as the units of observations and the variance is expressed in square of the units of the observations.

(ii) The standard deviation and the variance are zero if all the observations have some constant value. If 'C' is a constant, then

$$\text{S.D.}(C) = 0 \quad \text{and} \quad \text{Var}(C) = 0$$

(iii) The standard deviation and the variance do not change by change of origin.

Thus $\text{S.D.}(X + A) = \text{S.D.}(X)$ and $\text{Var}(X + A) = \text{Var}(X)$

Also $\text{S.D.}(X - A) = \text{S.D.}(X)$ and $\text{Var}(X - A) = \text{Var}(X)$

(iv) The standard deviation and the variance are affected by change of scale.

Thus S.D. $(bX) = |b|$ S.D.(X) and $\text{Var}(bX) = b^2 \text{Var}(X)$

and S.D.$\left(\frac{X}{b}\right) = \frac{1}{|b|}$ S.D.(X) and $\text{Var}\left(\frac{X}{b}\right) = \frac{1}{b^2}\text{Var}(X)$.

(v) If we have two sets of data with n_1 and n_2 observations having means $\bar{X}_1$ and $\bar{X}_2$ with variances S_1^2 and S_2^2, respectively, then the combined standard deviation, S_c of all the $n = n_1 + n_2$ observations is given by

$$S_c = \sqrt{\frac{n_1S_1^2 + n_2S_2^2}{n_1 + n_2} + \frac{n_1n_2}{(n_1+n_2)^2}(\bar{X}_1 - \bar{X}_2)^2}$$

and the combined variance S_c^2 is given by

$$S_c^2 = \frac{n_1S_1^2 + n_2S_2^2}{n_1 + n_2} + \frac{n_1n_2}{(n_1+n_2)^2}(\bar{X}_1 - \bar{X}_2)^2$$

(vi) If X and Y are independent random variables then,

$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ and $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$

Note: Independence of random variables has been discussed in Chapter 8. Students should know that if X and Y are independent random variables, then $E(XY) = E(X) \,.\, E(Y)$

Example 4.7

A manufacturer of flashlight batteries took a sample of 5 batteries from a day's production and used them continuously until they were drained. The number of hours they were used until failure were 342, 426,317, 545, 630. Compute the variance and coefficient of variation.

Solution:

The necessary calculations are given below:

X	$(X - \bar{X})$	$(X - \bar{X})^2$
342	− 110	12100
426	− 26	676
317	− 135	18225
545	+ 93	8649
630	+ 178	31684
$\Sigma X = 2260$	$\Sigma(X - \bar{X}) = 0$	$\Sigma(X - \bar{X})^2 = 71334$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{2260}{5} = 452$$

$$S^2 = \frac{\Sigma(X-\bar{X})^2}{n} = \frac{71334}{5} = 14266.8,\ S = \sqrt{14266.8} = 119.4437$$

$$C.V. = \frac{S}{\bar{X}} \times 100 = \frac{119.4437}{452} \times 100 = 26.43\ \%$$

Example 4.8.

(i) Calculate standard deviation and coefficient of variation in respect of the marks obtained by 10 students given as: 50, 55, 57, 49, 54, 61, 64, 59, 58, 56.

(ii) How would your results be affected if it is decided to increase the marks of each of the above students by 5.

Solution:

(i) The necessary calculations are given below:

X	50	55	57	49	54	61	64	59	58	56	Total
D = X – 56	– 6	– 1	1	– 7	– 2	5	8	3	2	0	3
D^2	36	1	1	49	4	25	64	9	4	0	193

$$\bar{X} = A + \frac{\Sigma D}{n} = 56 + \frac{3}{10} = 56 + 0.3 = 56.3$$

$$S = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2} = \sqrt{\frac{193}{10} - \left(\frac{3}{10}\right)^2} = \sqrt{19.3 - 0.09}$$

$$= \sqrt{19.21} = 4.383$$

$$\text{Coefficient of Variation (C.V.)} = \frac{S}{\bar{X}} \times 100 = \frac{4.383}{56.3} \times 100 = 7.78\ \%$$

(ii) When each item is increased by 5, the value of the arithmetic mean would be increased by 5 because when a constant is added to all the items, the value of arithmetic mean is increased by the constant added. Thus the arithmetic mean would be 56.3 + 5 = 61.3. The value of the standard deviation would not be affected when every item is increased by 5. However, the value of the coefficient of variation would be affected because mean is increased by 5.

$$\text{Coefficient of Variation (C.V.)} = \frac{S}{\bar{X}} \times 100 = \frac{4.383}{61.3} \times 100 = 7.15\ \%$$

Example 4.9.

The arithmetic mean and standard deviation of a series of 20 items were calculated by a student as 20 cm. and 5 cm. respectively. But while calculating them an item 13 was misread as 30. Find the correct values of arithmetic mean and standard deviation.

Solution:

$$\bar{X} = \frac{\Sigma X}{n} \quad \text{or} \quad \Sigma X = n\bar{X}$$

Incorrect $\bar{X} = 20$, Incorrect $\Sigma X = n\bar{X} = 20(20) = 400$

Correct ΣX = Incorrect ΣX – Wrong item + Correct item

$= 400 - 30 + 13 = 383$

$$\text{Correct } \bar{X} = \frac{\text{Correct } \Sigma X}{n} = \frac{383}{20} = 19.15$$

$$\text{We have } S^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 = \frac{\Sigma X^2}{n} - (\bar{X})^2 \text{ or } \Sigma X^2 = n[S^2 + (\bar{X})^2]$$

$$\text{Incorrect } \Sigma X^2 = 20[25 + 400] = 20(425) = 8500$$

$$\text{Correct } \Sigma X^2 = \text{Incorrect } \Sigma X^2 - (\text{Wrong item})^2 + (\text{Correct item})^2$$

$$= 8500 - (30)^2 + (13)^2 = 8500 - 900 + 169 = 7769$$

$$\text{Correct } S^2 = \frac{\text{Correct } \Sigma X^2}{n} - \left(\frac{\text{Correct } \Sigma X}{n}\right)^2 = \frac{7769}{20} - \left(\frac{383}{20}\right)^2 = 21.73$$

$$\text{Correct } S = \sqrt{21.73} = 4.66$$

Hence, the correct values of mean and standard deviation are 19.15 and 4.66 respectively.

Example 4.10.

(a) Find the mean, variance, standard deviation and coefficient of variation of the first **n** natural numbers 1, 2, 3, ..., n.

(b) Computer the mean, variance and coefficient of variation of the set of numbers 218, 209, 212, 211, 215 by using 208 as assumed mean.

(c) For a set of 10 numbers $\Sigma(X - \bar{X})^2 = 160$ and $\Sigma X^2 = 1160$. Find coefficient of variation.

(d) The time (in minutes) that a random sample of 12 employees took to complete a task is:

13, 49, 40, 28, 23, 15, 17, 21, 37, 66, 30, 45

Compute the mean and standard deviation for this sample. Also compute the percentage of observations lying within limits $\bar{X} \pm S$.

Solution:

(a) We have

$$\Sigma X = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

$$\Sigma X^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{n(n+1)/2}{n} = \frac{n+1}{2}$$

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2 = \frac{n(n+1)(2n+1)}{6n} - \left[\frac{n(n+1)}{2n}\right]^2$$

$$S^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n+1}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$S^2 = \frac{n+1}{2}\left[\frac{4n+2-3n-3}{6}\right] = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12} \quad \text{or} \quad S = \sqrt{\frac{n^2-1}{12}}$$

Coefficient of variation (C.V) $= \frac{S}{\bar{X}} \times 100$

$$= \frac{\sqrt{(n^2-1)/12}}{(n+1)/2} = \sqrt{\frac{(n+1)(n-1)/12}{(n+1)(n+1)/4}} = \sqrt{\frac{n-1}{3(n+1)}}$$

(b) The necessary calculations are given below:

Here A = 208

X	218	209	212	211	215	Total
D = X − 208	10	1	4	3	7	$\sum D = 25$
D^2	100	1	16	9	49	$\sum D^2 = 175$

$$\bar{X} = A + \frac{\sum D}{n} = 208 + \frac{25}{5} = 208 + 5 = 213$$

$$S^2 = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2 = \frac{175}{5} - \left(\frac{25}{5}\right)^2 = 35 - 25 = 10$$

Coefficient of variation (C.V) $= \frac{S}{\bar{X}} \times 100 = \frac{\sqrt{10}}{213} \times 100 = 1.5\%$

(c) Here, $n = 10$, $\sum(X - \bar{X})^2 = 160$ and $\sum X^2 = 1160$

$$S^2 = \frac{\sum(X-\bar{X})^2}{n} = \frac{160}{10} = 16 \qquad S = \sqrt{16} = 4$$

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2 = \frac{\sum X^2}{n} - (\bar{X})^2$$

$$16 = \frac{1160}{10} - (\bar{X})^2 \text{ or } 16 = 116 - (\bar{X})^2 \text{ or } (\bar{X})^2 = 116 - 16 = 100 \text{ or } \bar{X} = 10$$

$$C.V = \frac{S}{\bar{X}} \times 100 = \frac{4}{10} \times 100 = 40\%$$

(d) Here, $n = 12$, $\sum X = 384$ and $\sum X^2 = 15088$

$$\bar{X} = \frac{\sum X}{n} = \frac{384}{12} = 32$$

$$S = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2} = \sqrt{\frac{15088}{12} - \left(\frac{384}{12}\right)^2} = 15.28$$

$\bar{X} \pm S = 32 \pm 15.28 = 16.72, 47.28$

Number of observations lying within these limits = 8

Thus, proportion of observations = 8/12 = 0.67 or 67%

Example 4.11.

(i) Coefficient of variation of two series are 58 % and 69 %, their standard deviations are 21.2 and 15.6 respectively. What are their arithmetic means?

(ii) For a set of 100 observations, the sum of the deviations from 4 cm. is –11 cm. and the sum of the squares of these deviations is 257 cm. Find the coefficient of variation.

Solution:

(i) **For Series I:**

$$C.V. = \frac{S}{\bar{X}} \times 100$$

Here, C.V. = 58 % and S = 21.2

$$\text{Hence, } 58 = \frac{21.2}{\bar{X}} \times 100 \quad \text{or} \quad 58\bar{X} = 2120 \quad \text{or} \quad \bar{X} = \frac{2120}{58} = 36.55$$

For Series II:

$$C.V. = \frac{S}{\bar{X}} \times 100$$

Here, C.V. = 69 % and S = 15.6

$$\text{Hence, } 69 = \frac{15.6}{\bar{X}} \times 100 \quad \text{or} \quad 69\bar{X} = 1560 \quad \text{or} \quad \bar{X} = \frac{1560}{69} = 22.61$$

(ii) Here, $n = 100$, $A = 4$, $\Sigma D = -11$, $\Sigma D^2 = 257$

$$\bar{X} = A + \frac{\Sigma D}{n} = 4 + \frac{(-11)}{100} = 4 - 0.11 = 3.89$$

$$S = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2} = \sqrt{\frac{257}{100} - \left(\frac{-11}{100}\right)^2} = 1.6$$

$$C.V. = \frac{S}{\bar{X}} \times 100 = \frac{1.6}{3.89} \times 100 = 41.13\ \%$$

Example 4.12.

(i) After settlement the average daily wages in a factory had increased from Rs.120 to Rs.180 and the standard deviation had increased from Rs. 2.5 to Rs. 3.75. After settlement the wages have become higher and more uniform. Comment.

(ii) Examine whether the following data for obtaining the variance are consistent or not.

$$n = 120, \quad \Sigma X = -125, \quad \Sigma X^2 = 128.$$

(iii) In a surprise checking of passengers in a local bus, 20 passengers without tickets were caught. The sum of squares and the standard deviation of the amount found in their pockets were Rs.2000 and Rs.6 respectively. If the total fine imposed is equal to the amount discovered from them, and fine imposed is uniform. What is the amount each one of them will have to pay as fine?

Solution:

(i) Before settlement the average daily wages were Rs.120 with standard deviation as Rs.2.5. So coefficient of variation is,

$$C.V. = \frac{S}{\bar{X}} \times 100 = \frac{2.5}{120} \times 100 = 2.083\ \%$$

After settlement average daily wages become Rs.180 and standard deviation is Rs.3.75. So coefficient of variation is,

$$C.V. = \frac{S}{\bar{X}} \times 100 = \frac{3.75}{180} \times 100 = 2.083\ \%$$

Thus it is incorrect to say that the wages have become more uniform. Variability remains the same as the coefficient of variation remains the same.

(ii) We have,

$$S^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 = \frac{128}{120} - \left(-\frac{125}{120}\right)^2 = -0.0184$$

Variance can never be negative, therefore the data is inconsistent.

(iii) Here, $n = 20, \quad \Sigma X^2 = 2000, \quad S = 6, \quad S^2 = 36$

$$S^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 = \frac{\Sigma X^2}{n} - (\bar{X})^2$$

$$36 = \frac{2000}{20} - (\bar{X})^2 = 100 - (\bar{X})^2 \text{ or } (\bar{X})^2 = 100 - 36 = 64 \text{ or } \bar{X} = 8$$

Total amount recovered $= \Sigma X = n\bar{X} = 20(8) =$ Rs.160.

Hence, each of the passengers shall have to pay Rs.8 as fine.

Example 4.13

The scores obtained by five students on a set of examination papers were 64, 66, 68, 70, 72. Compute the variance of these scores. Without recalculation, deduce the variance of scores if these are changed by (i) adding 10 points to all scores (ii) increasing all scores by 10 % (iii) decreasing all scores by 10 %.

Solution:

The necessary calculations are given below:

X	64	66	68	70	72	$\Sigma X = 340$
X^2	4096	4356	4624	4900	5184	$\Sigma X^2 = 23160$

$$Var(X) = S^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 = \frac{23160}{5} - \left(\frac{340}{5}\right)^2 = 8$$

(i) By adding 10 points to all scores, we get $Y = X + 10$, then
$$Var(Y) = Var(X + 10) = Var(X) + Var(10) = 8 + 0 = 8$$
(ii) Increasing all scores by 10 %, we get $Y = 1.1\,X$, then
$$Var(Y) = Var(1.1\,X) = 1.21 Var(X) = 1.21(8) = 9.68$$
(iii) Decreasing all scores by 10 %, we get $Y = 0.9\,X$, then
$$Var\,(Y) = Var(0.9\,X) = 0.81\,Var(X) = 0.81(8) = 6.48$$

Example 4.14

An exercise physiologist measured the heart rates of 20 people who had been placed on a long distance running program. A frequency distribution for these rates is displayed below:

Rate (X)	67	68	69	70	72	75
Frequency (f)	1	1	3	5	8	2

Compute the sample standard deviation and coefficient of variation of the heart rates.

Solution:

The necessary calculations are given below:

X	f	fX	fX^2
67	1	67	4489
68	1	68	4624
69	3	207	14283
70	5	350	24500
72	8	576	41472
75	2	150	11250
Total	20	1418	100618

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{1418}{20} = 70.9$$

$$S = \sqrt{\frac{\Sigma fX^2}{\Sigma f} - \left(\frac{\Sigma fX}{\Sigma f}\right)^2} = \sqrt{\frac{100618}{20} - \left(\frac{1418}{20}\right)^2} = 2.02$$

$$C.V. = \frac{S}{\bar{X}} \times 100 = \frac{2.02}{70.9} \times 100 = 2.85\ \%$$

Example 4.15.

Calculate the standard deviation and coefficient of variation of the following distribution taking $D = X - 32.5$.

Age (years)	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
No. of persons	170	110	80	45	40	35

Solution:

The necessary calculations are given below:

Age (years)	X	f	D = X − 32.5	fD	fD^2
20 − 25	22.5	170	− 10	− 1700	17000
25 − 30	27.5	110	− 5	− 550	2750
30 − 35	32.5	80	0	0	0
35 − 40	37.5	45	5	225	1125
40 − 45	42.5	40	10	400	4000
45 − 50	47.5	35	15	525	7875
Total		480		− 1100	32750

$$\bar{X} = A + \frac{\Sigma fD}{\Sigma f} = 32.5 + \frac{(-1100)}{480} = 32.5 - 2.292 = 30.208$$

$$S = \sqrt{\frac{\Sigma fD^2}{\Sigma f} - \left(\frac{\Sigma fD}{\Sigma f}\right)^2} = \sqrt{\frac{32750}{480} - \left(\frac{-1100}{480}\right)^2} = 7.936$$

$$C.V. = \frac{S}{\bar{X}} \times 100 = \frac{7.936}{30.208} \times 100 = 26.27\,\%$$

Example 4.16

Compute the standard deviation using step deviation method of a sample of ages of restaurant workers who are insured under a group health insurance policy.

Ages	15 − 24	25 − 34	35 − 44	45 − 54	55 − 64
Frequency	22	27	41	35	25

Solution:

The necessary calculations are given below:

Ages	X	f	A = 39.5, h = 10 $u = \frac{X - 39.5}{10}$	fu	fu^2
15 − 24	19.5	22	− 2	− 44	88
25 − 34	29.5	27	− 1	− 27	27
35 − 44	39.5	41	0	0	0
45 − 54	49.5	35	+ 1	+ 35	35
55 − 64	59.5	25	+ 2	+ 50	100
Total		150		14	250

$$S = \sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2} \times h = \sqrt{\frac{250}{150} - \left(\frac{14}{150}\right)^2} \times 10 = 12.88$$

4.12. SHEPPARDS CORRECTIONS

In grouped data the different observations are put into the same class. In the calculation of variance or standard deviation for grouped data, the frequency f is multiplied with X which is the mid-point of the respective class. Thus it is assumed that all the observations in a class are centered at X. But this is not true because the observations are spread out in the class. This assumption introduces some error in the calculation of S^2 and S. The value of S^2 and S can be corrected to some extent by applying sheppards correction. Thus

$$S^2(\text{corrected}) = S^2 - \frac{h^2}{12}, \quad \text{where h is the uniform class interval}$$

$$S(\text{corrected}) = \sqrt{S^2 - \frac{h^2}{12}}$$

This correction is applied in grouped data which have almost equal tails in the start and at the end of the data. If data has a longer tail on any side, this correction is not applied. If size of the class interval 'h' is not the same in all the classes, the correction is not applicable.

4.12.1. CORRECTED COEFFICIENT OF VARIATION

When the corrected standard deviation is used in the calculation of the coefficient of variation, we get what is called the corrected coefficient of variation. Thus

$$\text{Corrected coefficient of variation} = \frac{S(\text{corrected})}{\bar{X}} \times 100$$

4.13. COMBINED VARIANCE

Like combined mean, the combined variance or standard deviation can be calculated for different sets of data. Suppose we have two sets of data containing n_1 and n_2 observations with means $\bar{X}_1$ and $\bar{X}_2$ and variances S_1^2 and S_2^2. If $\bar{X}_c$ is the combined mean and S_c^2 is the combined variance of $n_1 + n_2$ observations, then combined variance is given by

$$S_c^2 = \frac{n_1 S_1^2 + n_2 S_2^2 + n_1(\bar{X}_1 - \bar{X}\ )^2 + n_2(\bar{X}_2 - \bar{X}_c)^2}{n_1 + n_2}$$

It can also be written as $$S_c^2 = \frac{n_1\left[S_1^2 + (\bar{X}_1 - \bar{X}_c)^2\right] + n_2\left[S_2^2 + (\bar{X}_2 - \bar{X}_c)^2\right]}{n_1 + n_2}$$

where $$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

The combined standard deviation S_c can be calculated by taking the square root of S_c^2. The formula for combined variance can be generalized as:

$$S_c^2 = \frac{\Sigma n_i\left[S_i^2 + (\bar{X}_i - \bar{X}_c)^2\right]}{\Sigma n_i}$$ here $i = 1, 2, 3, \cdots, k$ and $\bar{X}_c = \frac{\Sigma n_i \bar{X}_i}{\Sigma n_i}$

Example 4.17.

(i) For a group of 50 male workers, the mean and standard deviation of their hourly wages are Rs. 63 and Rs. 9 respectively. For a group of 40 female workers, these values are Rs.54 and Rs.6 respectively. Find the mean and variance of the combined group of 90 workers.

(ii) A distribution consists of three components with frequencies 200, 250 and 300 having means 25, 10 and 15 and standard deviations 3, 4 and 5 respectively. Find the coefficient of variation of the combined distribution.

Solution: (i) Here, $n_1 = 50, \bar{X}_1 = 63, S_1 = 9, \quad S_1^2 = 81$

$$n_2 = 40, \bar{X}_2 = 54, S_2 = 6, \quad S_2^2 = 36$$

$$\text{Combined A.M.} = \bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{50(63) + 40(54)}{50 + 40} = \frac{5310}{90} = 59$$

$$\text{Combined Variance} = S_c^2 = \frac{n_1\left[S_1^2 + (\bar{X}_1 - \bar{X}_c)^2\right] + n_2\left[S_2^2 + (\bar{X}_2 - \bar{X}_c)^2\right]}{n_1 + n_2}$$

$$= \frac{50\,[81 + (63-59)^2] + 40\,[36 + (54-59)^2]}{50 + 40}$$

$$= \frac{4850 + 2440}{90} = \frac{7290}{90} = 81$$

(ii) Here, $n_1 = 200, \quad \bar{X}_1 = 25, S_1 = 3, \quad S_1^2 = 9$

$$n_2 = 250, \quad \bar{X}_2 = 10, S_2 = 4, \quad S_2^2 = 16$$

$$n_3 = 300, \quad \bar{X}_3 = 15, S_3 = 5, \quad S_3^2 = 25$$

$$\text{Combined A.M.} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1 + n_2 + n_3} = \frac{200(25)+250(10)+300(15)}{200 + 250 + 300} = \frac{12000}{750} = 16$$

Combined Variance

$$= \frac{n_1\left[S_1^2 + (\bar{X}_1 - \bar{X}_c)^2\right] + n_2\left[S_2^2 + (\bar{X}_2 - \bar{X}_c)^2\right] + n_3\left[S_3^2 + (\bar{X}_3 - \bar{X}_c)^2\right]}{n_1 + n_2 + n_3}$$

$$= \frac{200\,[9 + (25-16)^2] + 250\,[16 + (10-16)^2] + 300\,[25 + (15-16)^2]}{200 + 250 + 300}$$

$$= \frac{18000 + 13000 + 7800}{750} = \frac{38800}{750} = 51.733$$

$$\text{Combined standard deviation} = \sqrt{51.733} = 7.193$$

$$\text{Combined C.V.} = \frac{\text{Combined standard deviation}}{\text{Combined A.M.}} \times 100 = \frac{7.193}{16} \times 100 = 44.9$$

4.14. MOMENTS

Moments are defined as the mean of the different powers of the deviations of the observations taken from their mean. For sample data, the first four moments about the mean $\bar{X}$ are defined as

$$m_1 = \frac{\Sigma(X - \bar{X})}{n} = 0 \text{ (always)} \qquad m_2 = \frac{\Sigma(X - \bar{X})^2}{n}$$

$$m_3 = \frac{\Sigma(X - \bar{X})^3}{n} \qquad m_4 = \frac{\Sigma(X - \bar{X})^4}{n}.$$

For a frequency distribution the first four moments about the mean $\bar{X}$ are defined as

$$m_1 = \frac{\Sigma f(X - \bar{X})}{\Sigma f} = 0 \text{ (always)} \qquad m_2 = \frac{\Sigma f(X - \bar{X})^2}{\Sigma f}$$

$$m_3 = \frac{\Sigma f(X - \bar{X})^3}{\Sigma f} \qquad m_4 = \frac{\Sigma f(X - \bar{X})^4}{\Sigma f}.$$

The terms m_1, m_2, m_3 and m_4 are called **first, second, third and fourth** moments about the mean. They are also called the **central moments** or the **mean moments.** Value of the first moment m_1 is always equal to zero. The second moment $m_2 = S^2$ (Variance) and is always positive. The third moment m_3 can be negative, zero or positive. For a frequency distribution, when

$m_3 =$ negative, the distribution has negative skewness,

$m_3 =$ 0, the distribution is symmetrical,

and if $m_3 =$ positive, the distribution is positively skewed.

In general, the rth moment about the mean $\bar{X}$ is defined as

$$m_r = \frac{\Sigma f(X - \bar{X})^r}{\Sigma f}$$

If we put r = 0, we get $m_0 = \dfrac{\Sigma f(X - \bar{X})^0}{\Sigma f} = 1$. Thus m_0 is always equal to 1.

The moments about the mean are calculated to describe different characteristics of the distribution. Moments of the order higher than four can also be calculated but the second, third and fourth moments are the most important. With the help of these moments, we can find the spread and shape of the distribution.

For population data, the moments for a frequency distribution are defined as:

$$\mu_1 = \frac{\Sigma f(X - \mu)}{\Sigma f} = 0 \text{ (always)} \qquad \mu_2 = \frac{\Sigma f(X - \mu)^2}{\Sigma f} = \sigma^2$$

$$\mu_3 = \frac{\Sigma f(X - \mu)^3}{\Sigma f} \qquad \mu_4 = \frac{\Sigma f(X - \mu)^4}{\Sigma f} \text{ where } \mu = \frac{\Sigma fX}{\Sigma f}.$$

4.14.1. MOMENT RATIOS

Some ratios are calculated with the help of moments about the mean. For sample data, the important ratios are b_1 and b_2 where,

$$b_1 = \frac{m_3^2}{m_2^3} \quad \text{and} \quad b_2 = \frac{m_4}{m_2^2}$$

b_1 is called **moment coefficient of skewness** and b_2 is called **moment coefficient of kurtosis**, a term which will be explained later. The value of b_1 is either zero or positive. Its value cannot be negative. When $b_1 = 0$, the distribution is symmetrical when $b_1 > 0$, the distribution is either positively or negatively skewed.

Thus b_1 measures the **magnitude** of skewness. The direction of skewness is not determined by b_1. This direction is determined by the algebraic sign of m_3. The corresponding ratios for the population data are β_1 (beta one) and β_2 (beta two), where

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2}$$

4.14.2. RAW MOMENTS

The moments about the mean $\bar{X}$ are calculated by taking deviations of the observations from the mean $\bar{X}$. Sometimes the mean $\bar{X}$ is not a whole number or the mean contains many decimal places. Suppose $\bar{X} = \frac{\Sigma X}{n} = \frac{400}{6} = 66.66667$. When it is rounded off up to two places of decimals, we write $\bar{X} = 66.67$. Now this value of $\bar{X}$ is approximate and any calculation based on this value will have some error. If we write the value of $\bar{X}$, say, up to four places of decimals, the computational work requires a lot of time and labour. We can overcome this problem by calculating moments called raw moments. First we calculate the raw moments and then these raw moments are used to calculate the moments about the mean.

Raw moments are of different types. For sample dâta, the first four raw moments are denoted by m_1', m_2', m_3', and m_4', For population data, the first four raw moments are denoted by μ_1', μ_2', μ_3', and μ_4'.

4.14.3. RAW MOMENTS ABOUT THE ARBITRARY ORIGIN

Suppose we take some item of the given data and denote it by '*A*'. It is called an **arbitrary origin** or the **provisional mean**. The arbitrary origin is taken somewhere in the middle of the given data. The deviations of the observations are taken from the origin '*A*'. Let the deviations be denoted by D where $D = X - A$. The first four raw moments about the origin at $X = A$ are given by

$$m_1' = \frac{\Sigma f D}{\Sigma f} \qquad m_2' = \frac{\Sigma f D^2}{\Sigma f} \qquad m_3' = \frac{\Sigma f D^3}{\Sigma f} \qquad m_4' = \frac{\Sigma f D^4}{\Sigma f}$$

In general, the rth moment about the arbitrary origin is written as $m_r' = \frac{\Sigma f D^r}{\Sigma f}$

If we put $r = 0$, we get $m_0' = \frac{\Sigma f D^0}{\Sigma f} = 1$. Thus m_0' is always equal to 1.

4.14.4. RELATION BETWEEN THE MEAN MOMENTS AND THE RAW MOMENTS

The moments about the mean can be calculated from the raw moments about the arbitrary origin by using the following relations :

$$m_1 = 0$$

$$m_2 = m_2' - (m_1')^2$$

$$m_3 = m_3' - 3m_1' m_2' + 2(m_1')^3$$

$$m_4 = m_4' - 4m_1' m_3' + 6(m_1')^2 m_2' - 3(m_1')^4$$

4.14.5. RAW MOMENTS ABOUT X = 0

When the arbitrary origin 'A' is equal to zero, we have $X - A = X$. Thus the raw moments about $X = 0$ are given by

$$m_1' = \frac{\Sigma f X}{\Sigma f} = \bar{X} \qquad m_2' = \frac{\Sigma f X^2}{\Sigma f} \qquad m_3' = \frac{\Sigma f X^3}{\Sigma f} \qquad m_4' = \frac{\Sigma f X^4}{\Sigma f}$$

In general, the rth moment about the origin is defined as $m_r' = \dfrac{\Sigma f X^r}{\Sigma f}$.

When the magnitude of the given observations is small, it is convenient to calculate the set of moments called **raw moments about zero**. With the help of these moments, the central moments can be calculated by using the relations given earlier.

4.14.6. RAW MOMENTS IN CLASS INTERVAL UNITS

Let $u = \dfrac{X - A}{h}$, where 'A' is an arbitrary origin and 'h' is the class interval. The raw moments based on the variable 'u', are called the raw moments in the class interval units. These are defined as :

$$m_1' = \frac{\Sigma f u}{\Sigma f} \qquad m_2' = \frac{\Sigma f u^2}{\Sigma f} \qquad m_3' = \frac{\Sigma f u^3}{\Sigma f} \qquad m_4' = \frac{\Sigma f u^4}{\Sigma f}$$

In general, the rth moment in terms of class intervals units is $m_r' = \dfrac{\Sigma f u^r}{\Sigma f}$

It is very easy to calculate these moments. With the help of these moments, the moments about the mean can be calculated by using the relations :

$$m_1 = 0$$

$$m_2 = \left[m_2' - (m_1')^2\right]h^2$$

$$m_3 = \left[m_3' - 3m_1' m_2' + 2(m_1')^3\right]h^3$$

$$m_4 = \left[m_4' - 4m_1' m_3' + 6(m_1')^2 m_2' - 3(m_1')^4\right]h^4.$$

The moments about the mean can be calculated by using any set of raw moments. For grouped data, where class interval is uniform, the raw moments about class interval units are the easiest in calculations.

4.14.7 SHEPPARD'S CORRECTION

In case of the grouped data, the mid-points of the classes are calculated. These mid-points are denoted by X. In the calculation of the moments, it is assumed that the frequency of a certain class is concentrated on the centre or the mid-point of the class. This assumption is not correct. In fact, the frequency is distributed uniformly in the respective class. Because of this assumption, certain errors are introduced in the calculations of the central moments. Sheppard suggested certain corrections in the moments. The corrected central moments are given below:

$$m_2\,(\text{corrected}) = m_2 - \frac{h^2}{12}$$

$$m_3\,(\text{corrected}) = m_3$$

$$m_4\,(\text{corrected}) = m_4 - \frac{h^2}{2}m_2 + \frac{7}{240}h^4$$

where h is the uniform class interval of the grouped data. If the class interval is not uniform for all the groups, the Sheppard's corrections are not applicable. These corrections are also not applicable for highly skewed grouped data.

4.15. SYMMETRY

A distribution is called symmetrical if its frequency curve has the same shape on both sides of the central line which divides the curve into two equal halves. It has a hump (⌒) in the centre and the curve spreads equally on both sides. We say that it has equal tails on both sides. In a perfectly symmetrical distribution, the values of the mean, median and mode are equal and the two quartiles are at equal distance from the median. Fig.4.1.(a), Fig.4.1.(b) and Fig.4.1.(c) show symmetrical curves which have different standard deviations. In Fig.4.1.(a), the standard deviation is small, in Fig.4.1.(b), it is large and in Fig.4.1.(c), it has a moderate value.

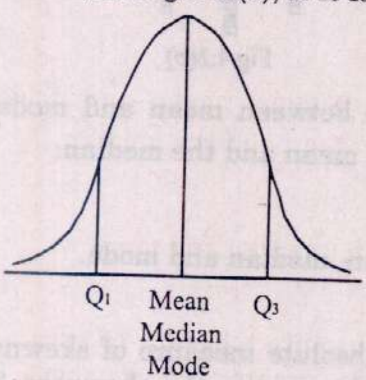


Fig. 4.1(a)

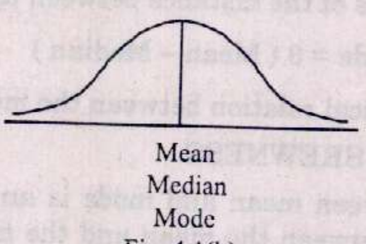


Fig. 4.1(b)

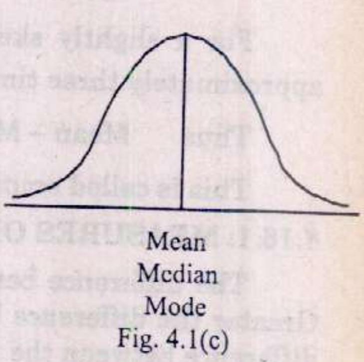


Fig. 4.1(c)

Following is a distribution which is symmetrical having the same values of the mean, median and mode.

Groups (marks of students)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	20	25	40	25	20

In this data, mean = median = mode = 25.

4.16. SKEWNESS

A distribution is called skewed if it is not symmetrical. A skewed distribution has a curve with a longer tail on any direction. If the longer tail is on the right side, the distribution is called positively skewed. If the longer tail is on the left side, the distribution is called negatively skewed. In case of positive skewness, the mean is greater than mode and for negative skewness, the mean is less than mode. In both cases the median lies between the mean and the mode. Fig.4.2.(a) shows positive and Fig.4.2.(b) shows negative skewness.

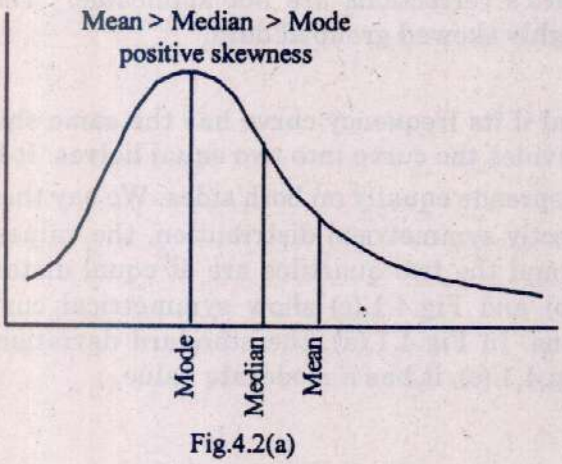


Fig.4.2(a)

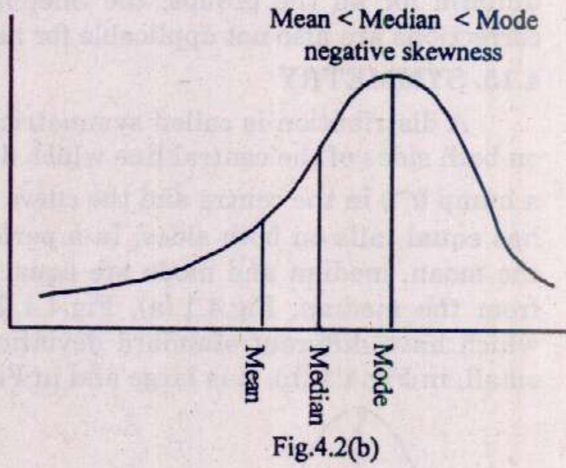


Fig.4.2(b)

For a slightly skewed distribution the distance between mean and mode is approximately three times of the distance between the mean and the median.

Thus Mean – Mode = 3 (Mean – Median)

This is called empirical relation between the mean, median and mode.

4.16.1. MEASURES OF SKEWNESS

The difference between mean and mode is an absolute measure of skewness. Greater the difference between the mean and the mode, greater the skewness. The difference between the mean and the mode is standardised and we get what is called relative measure of skewness or coefficient of skewness.

Karl Pearson suggested two measures of skewness which are defined below:

(i) Coefficient of skewness $= \dfrac{\text{Mean} - \text{Mode}}{S}$

(ii) Sometimes the mode is not clearly located in the data with the result that it cannot be calculated. In such cases the median is calculated and the formula used is:

Coefficient of skewness $= \dfrac{3(\text{Mean} - \text{Median})}{S}$

The above two formulas are designed in a manner such that the coefficient of skewness is zero for a symmetrical distribution, it is positive for a positively skewed distribution and negative for a negatively skewed distribution.

(iii) Another measure of skewness was suggested by Bowley, and after his name it is called the Bowley's coefficient of skewness. The formula is

$$\text{Bowley's Coefficient of skewness} = \frac{(Q_3 - \text{Median}) - (\text{Median} - Q_1)}{(Q_3 - \text{Median}) + (\text{Median} - Q_1)}$$

$$= \frac{Q_3 + Q_1 - 2\,\text{Median}}{Q_3 - Q_1}$$

Its value lies between -1 and $+1$.

This formula compares the difference between Q_3 and median and the difference between median and Q_1. If these two differences are equal, the distribution is symmetrical and the coefficient of skewness will be zero.

If$(Q_3 - \text{Median}) > (\text{Median} - Q_1)$, the distribution has positive skewness

If$(Q_3 - \text{Median}) < (\text{Median} - Q_1)$, the distribution has negative skewness.

In our country the distribution of agricultural land among farmers is positively skewed. It has a very long tail on the right side. There are very large number of farmers who has the land less than 5 acres and there are very small number of farmers who has the land more than 500 acres.

(iv) Karl Pearson has also given another measure of skewness based on moment ratios. The formula is

$$\text{Coefficient of Skewness} = \frac{\sqrt{\beta_1}\,(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

4.17. KURTOSIS

The term kurtosis is used for peakedness of the frequency curve. A curve which is tall and lean at the top is called leptokurtic. A curve which is flat like a plate at the top is called platykurtic. A curve with a moderate peak is called normal or mesokurtic. For a normal curve the value of $\beta_2 = 3$. When $\beta_2 > 3$, the curve is called leptokurtic. When $\beta_2 < 3$, the curve is called platykurtic. For a large sample data, the moment ratios $b_2 = \dfrac{m_4}{m_2^2}$ can be calculated. For a normal curve $b_2 = 3$. Kurtosis is a property of the large sample or a population. For small samples, b_2 should not be calculated.

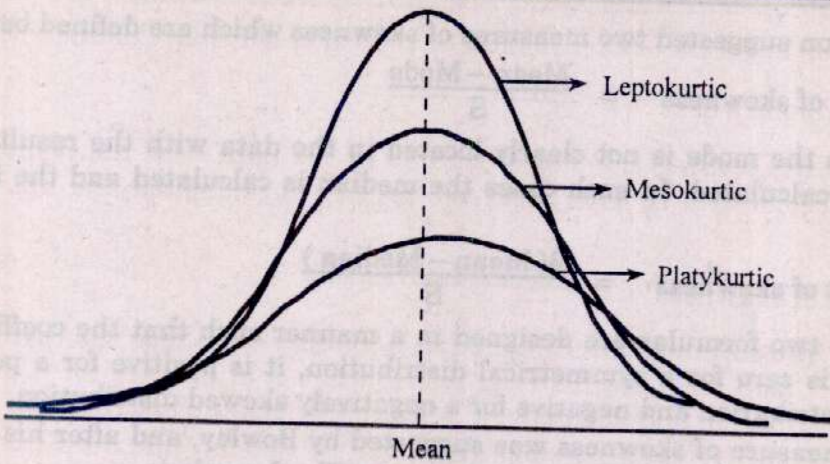


Fig. 4.3.

Example 4.18.

Calculate the first four moments about the mean for the following observations:

8, 10, 15, 22, 26, 30, 35, 40, 48.

Solution.

The necessary calculations are given below:

X	$X-\bar{X}$	$(X-\bar{X})^2$	$(X-\bar{X})^3$	$(X-\bar{X})^4$
8	−18	324	−5832	104976
10	−16	256	−4096	65536
15	−11	121	−1331	14641
22	−4	16	−64	256
26	0	0	0	0
30	4	16	64	256
35	9	81	729	6561
40	14	196	2744	38416
48	22	484	10648	234256
$\Sigma X = 234$	$\Sigma(X-\bar{X}) = 0$	$\Sigma(X-\bar{X})^2 = 1494$	$\Sigma(X-\bar{X})^3 = 2862$	$\Sigma(X-\bar{X})^4 = 464898$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{234}{9} = 26$$

Moments about the mean $\bar{X}$ are:

$$m_1 = \frac{\Sigma(X-\bar{X})}{n} = \frac{0}{9} = 0 \qquad m_2 = \frac{\Sigma(X-\bar{X})^2}{n} = \frac{1494}{9} = 166$$

$$m_3 = \frac{\Sigma(X-\bar{X})^3}{n} = \frac{2862}{9} = 318 \qquad m_4 = \frac{\Sigma(X-\bar{X})^4}{n} = \frac{464898}{9} = 51655.33$$

Example 4.19

Calculate the first four moments about the mean for the following data. Also calculate b_1 and b_2.

X	3	4	6	8	10
f	1	2	5	3	4

Solution:

The necessary calculations are given below:

X	f	fX	$(X-\bar{X})$	$f(X-\bar{X})$	$f(X-\bar{X})^2$	$f(X-\bar{X})^3$	$f(X-\bar{X})^4$
3	1	3	− 4	− 4	16	− 64	256
4	2	8	− 3	− 6	18	− 54	162
6	5	30	− 1	− 5	5	− 5	5
8	3	24	+ 1	+ 3	3	+ 3	3
10	4	40	+ 3	+ 12	36	+ 108	324
Total	15	105		0	78	−12	750

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{105}{15} = 7$$

Moments about the mean are

$$m_1 = \frac{\Sigma f(X-\bar{X})}{\Sigma f} = \frac{0}{15} = 0 \qquad m_2 = \frac{\Sigma f(X-\bar{X})^2}{\Sigma f} = \frac{78}{15} = 5.2$$

$$m_3 = \frac{\Sigma f(X-\bar{X})^3}{\Sigma f} = \frac{-12}{15} = -0.8 \qquad m_4 = \frac{\Sigma f(X-\bar{X})^4}{\Sigma f} = \frac{750}{15} = 50$$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(-0.8)^2}{(5.2)^3} = 0.005 \qquad b_2 = \frac{m_4}{m_2^2} = \frac{50}{(5.2)^2} = 1.85$$

Example 4.20

The following summations have been obtained by taking arbitrary origin $X = 9$.

$$\Sigma f = 165, \quad \Sigma fD = 216, \ \Sigma fD^2 = 1584, \ \Sigma fD^3 = 3168, \ \Sigma fD^4 = 33600$$

Compute the first four moments about the arbitrary origin $X = 9$ and convert them into the moments about the mean. Also calculate b_1 and b_2.

Solution:

Moments about $X = 9$ are:

$$m_1' = \frac{\Sigma fD}{\Sigma f} = \frac{216}{165} = 1.31 \qquad m_2' = \frac{\Sigma fD^2}{\Sigma f} = \frac{1584}{165} = 9.6$$

$$m_3' = \frac{\Sigma fD^3}{\Sigma f} = \frac{3168}{165} = 19.2 \qquad m_4' = \frac{\Sigma fD^4}{\Sigma f} = \frac{33600}{165} = 203.64$$

Moments about the mean are:

$$m_1 = m_1' - m_1' = 0$$

$$m_2 = m_2' - (m_1')^2 = 9.6 - (1.31)^2 = 7.88$$

$$m_3 = m_3' - 3m_1'm_2' + 2(m_1')^3 = 19.2 - 3(1.31)(9.6) + 2(1.31)^3 = -14.03$$

(since m_3 is negative, the distribution is negatively skewed).

$$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2 m_2' - 3(m_1')^4$$

$$= 203.64 - 4(1.31)(19.2) + 6(1.31)^2 (9.6) - 3(1.31)^4 = 193.04$$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(-14.03)^2}{(7.88)^3} = 0.40 \text{ and } b_2 = \frac{m_4}{m_2^2} = \frac{193.04}{(7.88)^2} = 3.11$$

Example 4.21

Given the following information:

$$\Sigma f = 290, \ \Sigma fX = 2610, \ \Sigma fX^2 = 23780, \ \Sigma fX^3 = 219530, \ \Sigma fX^4 = 2056100$$

Calculate first four moments about the arithmetic mean.

Solution:

Moments about $X = 0$ are:

$$m_1' = \frac{\Sigma fX}{\Sigma f} = \frac{2610}{290} = 9 \qquad m_2' = \frac{\Sigma fX^2}{\Sigma f} = \frac{23780}{290} = 82$$

$$m_3' = \frac{\Sigma fX^3}{\Sigma f} = \frac{219530}{290} = 757 \qquad m_4' = \frac{\Sigma fX^4}{\Sigma f} = \frac{2056100}{290} = 7090$$

Moments about the arithmetic mean are:

$m_1 = m_1' - m_1' = 0$

$m_2 = m_2' - (m_1')^2 = 82 - (9)^2 = 1$

$m_3 = m_3' - 3m_1'm_2' + 2(m_1')^3 = 757 - 3(9)(82) + 2(9)^3 = 757 - 2214 + 1458 = 1$

$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2 m_2' - 3(m_1')^4$

$= 7090 - 4(9)(757) + 6(9)^2(82) - 3(9)^4 = 7090 - 27252 + 39852 - 19683 = 7$

Example 4.22.

A survey was conducted by a manufacturing company to enquire the maximum price at which persons would be willing to buy their product. The following table gives the stated price (in rupees) by persons.

Price (in Rs.)	80 – 90	90 – 100	100 – 110	110 – 120	120 – 130
No. of persons	11	29	18	27	15

Calculate Karl Pearson's coefficient of skewness and interpret its value.

Solution:

The necessary calculations are given below:

Price (in Rs.)	X	f	$A = 105, h = 10$ $u = \left(\frac{X - 105}{10}\right)$	fu	fu^2
80 – 90	85	11	– 2	– 22	44
90 – 100	95	29	– 1	– 29	29
100 – 110	105	18	0	0	0
110 – 120	115	27	1	27	27
120 – 130	125	15	2	30	60
Total		100		6	160

$$\text{Mean} = A + \frac{\Sigma fu}{\Sigma f} \times h = 105 + \frac{6}{100} \times 10 = 105 + 0.6 = 105.6$$

$$\text{Mode} = l + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

where $l = 90$, $f_m = 29$, $f_1 = 11$, $f_2 = 18$ and $h = 10$

$$\text{Thus, Mode} = 90 + \frac{(29 - 11)}{(29 - 11) + (29 - 18)} \times 10 = 90 + 6.21 = 96.21$$

$$S = \sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2} \times h = \sqrt{\frac{160}{100} - \left(\frac{6}{100}\right)^2} \times 10 = 12.635$$

$$\text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{S} = \frac{105.6 - 96.21}{12.635} = 0.743$$

There is a high degree of positive skewness.

Example 4.23

Compute the Bowley's coefficient of skewness and interpret its value for the data given below:

Groups	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
Frequency	12	15	16	15	12

Solution:

The necessary calculations are given below:

Groups	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
f	12	15	16	15	12
c.f.	12	27	43	58	70

$$\text{Median} = \text{Value of}\left(\frac{n}{2}\right)\text{th item} = \text{Value of}\left(\frac{70}{2}\right)\text{th item} = 35\text{th item}$$

Median lies in the class 60 – 70.

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - c\right) = 60 + \frac{10}{16}\ (35 - 27) = 60 + 5 = 65$$

$$Q_1 = \text{Value of}\left(\frac{n}{4}\right)\text{th item} = \text{Value of}\left(\frac{70}{4}\right)\text{th item} = 17.5\text{th item}$$

Q_1 lies in the class 50 – 60.

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right) = 50 + \frac{10}{15}\ (17.5 - 12) = 50 + 3.67 = 53.67$$

$$Q_3 = \text{Value of}\left(\frac{3n}{4}\right)\text{th item} = \text{Value of}\left(\frac{3 \times 70}{4}\right)\text{th item} = 52.5\text{th item}$$

Q_3 lies in the class 70 – 80.

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right) = 70 + \frac{10}{15}\ (52.5 - 43) = 70 + 6.33 = 76.33$$

$$\text{Bowley's coefficient of skewness} = \frac{Q_3 + Q_1 - 2\text{ median}}{Q_3 - Q_1}$$

$$= \frac{76.33 + 53.67 - 2(65)}{76.33 - 53.67} = 0$$

Hence the distribution is symmetrical.

Example 4.24.

(i) In a frequency distribution the coefficient of skewness is 0.6. If the sum of the upper and the lower quartiles is 100 and the median is 38, find the value of the upper quartile.

(ii) For a moderately skewed distribution the mean price is Rs.20 and the median price is Rs.17. If the coefficient of variation is 20 %, find the coefficient of skewness of the distribution.

(iii) In a certain distribution the following results were obtained: Mean = 45, Median = 48, Coefficient of skewness = – 0.4. Find the value of the standard deviation.

(iv) Coefficient of skewness of a distribution is 0.64, its mean is 82 and mode is 50. Find the coefficient of variation.

Solution:

(i) $$\text{Coefficient of skewness} = \frac{Q_3 + Q_1 - 2\,\text{Median}}{Q_3 - Q_1}$$

Here, Coefficient of skewness = 0.6, $Q_3 + Q_1 = 100$, Median = 38, $Q_1 = 100 - Q_3$

$$\text{Therefore,}\quad 0.6 = \frac{100 - 2(38)}{Q_3 - 100 + Q_3} = \frac{24}{2\,Q_3 - 100}$$

$$0.6(2Q_3 - 100) = 24 \quad\text{or}\quad 1.2\,Q_3 - 60 = 24 \quad\text{or}\quad 1.2\,Q_3 = 24 + 60 = 84 \quad\text{or}$$

$$Q_3 = \frac{84}{1.2} = 70$$

(ii) $$\text{Coefficient of variation (C.V.)} = \frac{S}{\text{Mean}} \times 100$$

Here, Mean = 20, Median = 17, Coefficient of Variation = 20 %

$$\text{Therefore,}\quad 20 = \frac{S}{20} \times 100 \quad\text{or}\quad 5\,S = 20 \quad\text{or}\quad S = \frac{20}{5} = 4$$

$$\text{Coefficient of skewness} = \frac{3(\text{Mean} - \text{Median})}{S} = \frac{3(20 - 17)}{4} = 2.25$$

(iii) $$\text{Coefficient of skewness} = \frac{3(\text{Mean} - \text{Median})}{S}$$

Here, Mean = 45, Median = 48, Coefficient of skewness = – 0.4

$$\text{Therefore,}\quad -0.4 = \frac{3(45 - 48)}{S} = -\frac{9}{S} \quad\text{or}\quad S = \frac{9}{0.4} = 22.5$$

(iv) $$\text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{S}$$

Here, Mean = 82, Mode = 50, Coefficient of skewness = 0.64

$$\text{Therefore,}\quad 0.64 = \frac{82 - 50}{S} = \frac{32}{S} \quad\text{or}\quad S = \frac{32}{0.64} = 50$$

$$\text{Coefficient of Variation (C.V.)} = \frac{S}{\bar{X}} \times 100 = \frac{50}{82} \times 100 = 60.98\ \%$$

SHORT DEFINITIONS

Zain

Dispersion or Variation

The degree to which a set of values vary about their mean.

or

The spread or variability of scores about the measure of central tendency.

Measure of Dispersion

A measure of dispersion indicates how close or far apart the values are from the mean or other measure of central tendency.

or

A measure of dispersion describes the spread or scatter of the individual values around the central position.

Absolute Measure of Dispersion

An absolute measure of dispersion is that which measures the variation present between the observations in the unit of the variable.

Relative Measure of Dispersion

A relative measure of dispersion is that which measures the variation present between the observations relative to their average. It is used to compare the variation between two or more sets of data.

Range

The range is the difference between the largest and the smallest value in a set of observations.

or

The range of a data set is equal to the largest measurement minus the smallest measurement.

Interquartile Range

The interquartile range is obtained by subtracting the score of the first quartile from the score of the third quartile.

or

The interquartile range is the distance between the lower and upper quartiles.

Quartile Deviation or Semi-Interquartile Range

Quartile deviation is half of the difference between upper and lower quartiles.

or

Interquartile range divided by 2, is called the quartile deviation.

Mean Deviation

The arithmetic mean of the absolute values of the deviations from the arithmetic mean is known as mean deviation.

or

Sum of the deviation of each score from the mean, without regard to sign, divided by the number of scores is called mean deviation.

Variance

The variance is defined as the average squared deviation from the mean.

or

Sum of the squared deviations from the mean, divided by n, is called variance.

Standard Deviation

The positive square root of the average of the squared deviations from the mean is called standard deviation.

or

The positive square root of the sum of the squared deviations from the mean divided by n, is called standard deviation.

Coefficient of Variation

Coefficient of variation is defined as the standard deviation of a distribution divided by the mean of the distribution and multiplied by 100.

or

The ratio of the standard deviation to the arithmetic mean expressed as a percentage is known as coefficient of variation.

Empirical Rule

A rule that states the percentages of items that are within one, two, and three standard deviations from the mean for mound-shaped, or bell-shaped, distributions.

Symmetrical Distribution

A distribution is called symmetrical if its frequency curve has the same shape on both sides of the central line which divides the curve into two equal parts. In a perfectly symmetrical distribution, the values of the mean, median and mode are equal and the two quartiles are at equal distance from the median.

Skewed Distribution

A distribution in which the observations are concentrated at one end (tail) of the distribution, so that the most typical scores are relatively low (skewed to the right) or relatively high (skewed to the left).

Leptokurtic Distribution

A bell-shaped distribution characterized by a piling up of scores in the center of the distribution.

Mesokurtic Distribution

Bell-shaped distribution; "ideal" form of normal curve.

Platykurtic Distribution

A frequency distribution characterized by a flattening in the central position.

MULTIPLE – CHOICE QUESTIONS

1. The scatter in a series of values about the average is called:
   (a) central tendency (b) dispersion
   (c) skewness (d) symmetry

2. The measurements of spread or scatter of the individual values around the central position is called:
   (a) measures of dispersion (b) measures of central tendency
   (c) measures of skewness (d) measures of kurtosis
3. The measures used to calculate the variation present among the observations in the unit of the variable is called:
   (a) relative measures of dispersion (b) coefficient of skewness
   (c) absolute measures of dispersion (d) coefficient of variation
4. The measures used to calculate the variation present among the observations relative to their average is called:
   (a) coefficient of kurtosis (b) absolute measures of dispersion
   (c) quartile deviation (d) relative measures of dispersion
5. The degree to which numerical data tend to spread about an average value is called:
   (a) constant (b) flatness
   (c) variation (d) skewness
6. The measures of dispersion can never be:
   (a) positive (b) zero
   (c) negative (d) equal to 2
7. If all the scores on examination cluster around the mean, the dispersion is said to be:
   (a) small (b) large
   (c) normal (d) symmetrical
8. If there are many extreme scores on an examination, the dispersion is:
   (a) large (b) small
   (c) normal (d) symmetrical
9. Given below the four sets of observations. Which set has the minimum variation?
   (a) 46, 48, 50, 52, 54 (b) 30, 40, 50, 60, 70
   (c) 40, 50, 60, 70, 80 (d) 48, 49, 50, 51, 52.
10. Which of the following is an absolute measures of dispersion?
   (a) Coefficient of variation (b) Coefficient of dispersion
   (c) Standard deviation (d) Coefficient of skewness
11. The measure of dispersion which utilizes only extreme values is called:
   (a) mean (b) median
   (c) range (d) coefficient of variation
12. The measure of dispersion which uses only two observations is called:
   (a) range (b) quartile deviation
   (c) mean deviation (d) standard deviation
13. In quality control of manufactured items, the most commonly used measure of dispersion is:
   (a) range (b) average deviation
   (c) standard deviation (d) quartile deviation

14. The range of the scores 29, 3, 143, 27, 99 is:
(a) 140 (b) 143
(c) 146 (d) 70

15. If the observations of a variable X are – 4, – 20, – 30, – 44 and – 36, then the value of the range will be:
(a) – 48 (b) 40
(c) – 40 (d) 48

16. The range of the values – 5, – 8, – 10, 0, 6, 10 is:
(a) 0 (b) 10
(c) – 10 (d) 20

17. If $Y = aX \pm b$, where a and b are any two numbers and $a \neq 0$, then the range of Y values will be:
(a) range (X) (b) a range (X) + b
(c) a range (X) – b (d) $|a|$ range (X)

18. If the maximum value in a series is 25 and its range is 15, the minimum value of the series is:
(a) 10 (b) 15
(c) 25 (d) 35

19. Half of the difference between upper and lower quartiles is called:
(a) interquartile range (b) quartile deviation
(c) mean deviation (d) standard deviation

20. If $Q_3 = 20$ and $Q_1 = 10$, the coefficient of quartile deviation is:
(a) 3 (b) 1/3
(c) 2/3 (d) 1

21. Which measure of dispersion can be computed in case of open-end classes?
(a) standard deviation (b) range
(c) quartile deviation (d) coefficient of variation

22. If $Y = aX \pm b$, where a and b are any two constants and $a \neq 0$, then the quartile deviation of Y values is equal to:
(a) a Q.D(X) + b (b) $|a|$ Q.D(X)
(c) Q.D(X) – b (d) $|b|$ Q.D(X)

23. The sum of absolute deviations is minimum if these deviations are taken from the:
(a) mean (b) mode
(c) media. (d) upper quartile

24. The mean deviation is minimum when deviations are taken from:
(a) mean (b) mode
(c) median (d) zero

25. If $Y = aX \pm b$, where a and b are any two numbers but $a \neq 0$, then M.D(Y) is equal to:
(a) M.D(X) (b) M.D(X) ± b
(c) $|a|$ M.D(X) (d) M.D(Y) + M.D(X)

26. The mean deviation of the scores 12, 15, 18 is:
(a) 6 (b) 0
(c) 3 (d) 2

27. Mean deviation computed from a set of data is always:
(a) negative (b) equal to standard deviation
(c) more than standard deviation (d) less than standard deviation

28. The average of squared deviations from mean is called:
(a) mean deviation (b) variance
(c) standard deviation (d) coefficient of variation

29. The sum of squares of the deviations is minimum, when deviations are taken from:
(a) mean (b) mode
(c) median (d) zero

30. Which of the following measures of dispersion is expressed in the same units as the units of observation?
(a) Variance (b) Standard deviation
(c) Coefficient of variation (d) Coefficient of standard deviation

31. Which measure of dispersion has a different unit other than the unit of measurement of values:
(a) range (b) standard deviation
(c) variance (d) mean deviation

32. Which of the following is a unit free quantity:
(a) range (b) standard deviation
(c) coefficient of variation (d) arithmetic mean

33. If the dispersion is small, the standard deviation is:
(a) large (b) zero
(c) small (d) negative

34. The value of standard deviation changes by a change of:
(a) origin (b) scale
(c) algebraic signs (d) none

35. The standard deviation of a distribution divided by the mean of the distribution and expressing in percentage is called:
(a) coefficient of standard deviation (b) coefficient of skewness
(c) coefficient of quartile deviation (d) coefficient of variation

36. The positive square root of the mean of the squares of the deviations of observations from their mean is called:
(a) variance (b) range
(c) standard deviation (d) coefficient of variation

37. The variance is zero only if all observations are the:
(a) different (b) square
(c) square root (d) same

38. The standard deviation is independent of:

(a) change of origin (b) change of scale of measurement

(c) change of origin and scale of measurement (d) difficult to tell

39. If there are ten values each equal to 10, then standard deviation of these values is:

(a) 100 (b) 20

(c) 10 (d) 0

40. If X and Y are independent random variables, then $S.D.(X \pm Y)$ is equal to:

(a) $S.D(X) \pm S.D(Y)$ (b) $Var(X) \pm Var(Y)$

(c) $\sqrt{Var(X) \pm Var(Y)}$ (d) $\sqrt{Var(X) + Var(Y)}$

41. $S.D.(X) = 6$ and $S.D.(Y) = 8$. If X and Y are independent random variables, then $S.D.(X - Y)$ is:

(a) 2 (b) 10

(c) 14 (d) 100

42. For two independent variables X and Y if $S.D.(X) = 1$ and $S.D.(Y) = 3$, then $Var(3X - Y)$ is equal to:

(a) 0 (b) 6

(c) 18 (d) 12

43. If $Y = aX \pm b$, where a and b are any two constants and $a \neq 0$, then Var (Y) is equal to:

(a) $a\,Var(X)$ (b) $a\,Var(X) + b$

(c) $a^2 Var(X) - b$ (d) $a^2\,Var(X)$

44. If $Y = aX + b$, where a and b are any two numbers but $a \neq 0$, then $S.D.(Y)$ is equal to:

(a) $S.D(X)$ (b) $a.S.D(X)$

(c) $|a|S.D(X)$ (d) $a.S.D(X) + b$

45. The ratio of the standard deviation to the arithmetic mean expressed as a percentage is called:

(a) coefficient of standard deviation (b) coefficient of skewness

(c) coefficient of kurtosis (d) coefficient of variation

46. Which of the following statements is correct:

(a) The standard deviation of a constant is equal to unity

(b) The sum of absolute deviations is minimum if these deviations are taken from the mean.

(c) The second moment about origin equals variance

(d) The variance is positive quantity and is expressed in square of the units of the observations.

47. Which of the following statements is false?
   (a) The standard deviation is independent of change of origin
   (b) If the moment coefficient of kurtosis $\beta_2 = 3$, the distribution is meokurtic or normal.
   (c) If the frequency curve has the same shape on both sides of the central line which divides the curve into two equal parts, is called a symmetrical distribution.
   (d) Variance of the sum or difference of any two variables is equal to the sum, of their respective variances.

48. If $Var(X) = 25$, then $S.D\left(\frac{2X+5}{2}\right)$ is equal to:
   (a) 15/2 (b) 50
   (c) 25 (d) 5

49. To compare the variation of two or more than two series, we use
   (a) combined standard deviation (b) corrected standard deviation
   (c) coefficient of variation (d) coefficient of skewness

50. The standard deviation of – 5, – 5, – 5, – 5, – 5 is:
   (a) – 5 (b) + 5
   (c) 0 (d) – 25

51. Standard deviation is always calculated from:
   (a) mean (b) median
   (c) mode (d) lower quartile

52. The mean of an examination is 69, the median is 68, the mode is 67, and the standard deviation is 3. The measures of variation for this examination is:
   (a) 67 (b) 68
   (c) 69 (d) 3

53. The variance of 19, 21, 23, 25 and 27 is 8. The variance of 14, 16, 18, 20 and 22 is:
   (a) greater than 8. (b) 8
   (c) less than 8 (d) $8 - 5 = 3$

54. In a set of observations the variance is 50. All the observations are increased by 100 %. The variance of the increased observations will become:
   (a) 50 (b) 200
   (c) 100 (d) no change.

55. Three factories A, B, C have 100, 200 and 300 workers respectively. The mean of the wages is the same in the three factories. Which of the following statements is true?
   (a) There is greater variation in factory C.
   (b) Standard deviation in factory A is the smallest.
   (c) Standard deviation in all the three factories are equal.
   (d) None of the above.

56. An automobile manufacturer obtains data concerning the sales of six of its deals in the last week of 1996. The results indicate that the standard deviation of their sales equals $\hat{6}$ autos. If this is so, the variance of their sales equals:

(a) $\sqrt{6}$ (b) 6

(c) $\frac{1}{\sqrt{6}}$ (d) 36

57. If standard deviation of the values 2, 4, 6, 8 is 2.236, then standard deviation of the values 4, 8, 12, 16 is:

(a) 0 (b) 4.472 (c) 4.236 (d) 2.236

58. $\text{Var}(X) = 4$ and $\text{Var}(Y) = 9$. If X and Y are independent random variable then $\text{Var}(2X + Y)$ is:

(a) 13 (b) 17 (c) 25 (d) -1

59. If $\overline{X}$ = Rs. 20, S = Rs. 10, then coefficient of variation is:

(a) 45 % (b) 50 % (c) 60 % (d) 65 %

60. Which of the following measures of dispersion is independent of the units employed?

(a) Coefficient of variation (b) Quartile deviation

(c) Standard deviation (d) Range

61. In Sheppard's correction μ_2 is equal to:

(a) $\mu_2 + \frac{h^2}{12}$ (b) $\mu_2 - \frac{h^2}{12}$

(c) $\mu_2 - \frac{h}{12}$ (d) $\mu_2 - \frac{h^2}{2}$

62. The moments about mean are called:

(a) raw moments (b) central moments

(c) moments about origin (d) all of the above

63. The moments about origin are called:

(a) moments about zero (b) raw moments

(c) both (a) and (b) (d) neither (a) nor (b)

64. All odd order moments about mean in a symmetrical distribution are:

(a) positive (b) negative

(c) zero (d) three

65. The second moment about arithmetic mean is 16, the standard deviation will be:

(a) 16 (b) 4 (c) 2 (d) 0

66. The first and second moments about arbitrary constant are -2 and 13 respectively. The standard deviation will be:

(a) -2 (b) 3

(c) 9 (d) 13

67. Moment ratios β_1 and β_2 are:
    (a) independent of origin and scale of measurement
    (b) expressed in original unit of the data
    (c) unitless quantities (d) both (a) and (c)
68. The first moment about $X = 0$ of a distribution is 12.08. The mean is:
    (a) 10.80 (b) 10.08
    (c) 12.08 (d) 12.88
69. First two moments about the value 2 of a variable are 1 and 16. The variance will be:
    (a) 13 (b) 15
    (c) 16 (d) difficult to tell.
70. The first three moments of a distribution about the mean $\bar{X}$ are 0, 4 and 0. The distribution is:
    (a) symmetrical. (b) skewed to the left.
    (c) skewed to the right. (d) normal
71. If the third central moment is negative, the distribution will be:
    (a) symmetrical (b) positively skewed
    (c) negatively skewed (d) normal
72. If the third moment about mean is zero, then the distribution is:
    (a) positively skewed (b) negatively skewed
    (c) symmetrical (d) mesokurtic
73. Departure from symmetry is called:
    (a) second moment (b) kurtosis
    (c) skewness (d) variation
74. In a symmetrical distribution, the coefficient of skewness will be:
    (a) 0 (b) Q_1
    (c) Q_3 (d) 1
75. The lack of uniformity or symmetry is called:
    (a) skewness (b) dispersion
    (c) kurtosis (d) standard deviation
76. For a positively skewed distribution, mean is always:
    (a) less than the median. (b) less than the mode.
    (c) greater than the mode. (d) difficult to tell.
77. For a symmetrical distribution:
    (a) $\beta_1 > 0$ (b) $\beta_1 < 0$
    (c) $\beta_1 = 0$. (d) $\beta_1 = 3$.
78. If mean = 50, mode = 40 and standard deviation = 5, the distribution is:
    (a) positively skewed (b) negatively skewed
    (c) symmetrical (d) difficult to tell

79. If mean = 25, median = 30 and standard deviation = 15, the distribution will be:
(a) symmetrical (b) positively skewed
(c) negatively skewed (d) normal

80. If mean = 20, mode = 16 and standard deviation = 2, then coefficient of skewness is:
(a) 1 (b) 2 (c) 4 (d) – 2

81. If mean = 10, median = 8 and standard deviation = 6, then coefficient of skewness is:
(a) 1 (b) – 1 (c) 2/6 (d) 2

82. If the sum of deviations from median is not zero, then a distribution will be:
(a) symmetrical (b) skewed
(c) normal (d) all of the above

83. In case of positively skewed distribution, the extreme values lie in the:
(a) middle (b) left tail
(c) right tail (d) anywhere

84. Bowley's coefficient of skewness lies between:
(a) 0 and 1 (b) – 1 and + 1
(c) – 1 and 0 (d) – 2 and + 2

85. In a symmetrical distribution, $Q_3 - Q_1 = 20$, median = 15. Q_3 is equal to:
(a) 5 (b) 15 (c) 20 (d) 25

86. Which of the following is correct in a negatively skewed distribution?
(a) The arithmetic mean is greater than the mode.
(b) The arithmetic mean is greater than the median.
(c) (Q_3 – median) = (median – Q_1). (d) (Q_3 – median) < (median – Q_1).

87. The lower and upper quartiles of a distribution are 80 and 120 respectively, while median is 100. The shape of the distribution is:
(a) positively skewed. (b) negatively skewed.
(c) symmetrical. (d) normal

88. In a symmetrical distribution $Q_1 = 20$ and median = 30. The value of Q_3 is:
(a) 50 (b) 35 (c) 40 (d) 25

89. The degree of peakedness or flatness of a unimodal distribution is called:
(a) skewness (b) symmetry
(c) dispersion (d) kurtosis

90. For a leptokurtic distribution, the relation between second and fourth central moment is:
(a) $\mu_4 = 3\mu_2^2$ (b) $\mu_4 < 3\mu_2^2$
(c) $\mu_4 > 3\mu_2^2$ (d) $\mu_4 = \mu_2^2$

91. For a platykurtic distribution, the relation between μ_4 and μ_2 is:
(a) $\mu_4 < 3\mu_2^2$ (b) $3\mu_2^2 > \mu_4$
(c) both (a) and (b) (d) $\mu_4 = \mu_2$

92. For a mesokurtic distribution, the relation between fourth and second mean moment is:

(a) $\mu_4 = 3\mu_2^2$ (b) $\mu_4 > 3\mu_2^2$

(c) $\mu_4 < 3\mu_2^2$ (d) μ_4 / μ_2^2

93. The second and fourth moments about mean are 4 and 48 respectively, then the distribution is:

(a) leptokurtic (b) platykurtic

(c) mesokurtic or normal (d) positively skewed

94. In a mesokurtic or normal distribution, $\mu_4 = 243$. The standard deviation is:

(a) 81 (b) 27

(c) 9 (d) 3

95. The value of β_2 can be:

(a) less than 3 (b) greater than 3

(c) equal to 3 (d) all of the above

96. In a normal (mesokurtic) distribution:

(a) $\beta_1 = 0$ and $\beta_2 = 3$ (b) $\beta_1 = 3$ and $\beta_2 = 0$

(c) $\beta_1 = 0$ and $\beta_2 > 3$ (d) $\beta_1 = 0$ and $\beta_2 < 3$

97. Any frequency distribution, the following empirical relation holds:

(a) Quartile deviation $= \frac{2}{3}$ Standard deviation

(b) Mean deviation $= \frac{4}{5}$ Standard deviation

(c) Standard deviation $= \frac{5}{4}$ Mean deviation $= \frac{3}{2}$ Quartile deviation

(d) All of the above

Answers

1.	(b)	2.	(a)	3.	(c)	4.	(d)	5.	(c)	6.	(c)	7.	(a)	8.	(a)
9.	(d)	10.	(c)	11.	(c)	12.	(a)	13.	(a)	14.	(a)	15.	(b)	16.	(d)
17.	[illegible]	18.	(a)	19.	(b)	20.	(b)	21.	(c)	22.	(b)	23.	(c)	24.	(c)
25.	(c)	26.	(d)	27.	(d)	28.	(b)	29.	(a)	30.	(b)	31.	(c)	32.	(c)
33.	(c)	34.	(b)	35.	(d)	36.	(c)	37.	(d)	38.	(a)	39.	(d)	40.	(d)
41.	(b)	42.	(c)	43.	(d)	44.	(c)	45.	(d)	46.	(d)	47.	(d)	48.	(d)
49.	(c)	50.	(c)	51.	(a)	52.	(d)	53.	(b)	54.	(b)	55.	(d)	56.	(d)
57.	(b)	58.	(c)	59.	(b)	60.	(a)	61.	(b)	62.	(b)	63.	(c)	64.	(c)
65.	(b)	66.	(b)	67.	(c)	68.	(c)	69.	(b)	70.	(a)	71.	(c)	72.	(c)
73	(c)	74.	(a)	75.	(a)	76.	(c)	77.	(c)	78.	(a)	79.	(c)	80.	(b)
81.	(a)	82.	(b)	83.	(c)	84.	(b)	85.	(d)	86.	(d)	87.	(c)	88.	(c)
89.	(d)	90.	(c)	91.	(c)	92.	(a)	93.	(c)	94.	(d)	95.	(d)	96.	(a)
97.	(d)														

SHORT QUESTIONS

Q.1 Explain the meaning of the term dispersion or variation.
Q.2 What are the qualities of a good measure of dispersion?
Q.3 Write down the types of dispersion.
Q.4 Explain the absolute measure of dispersion.
Q.5 Define relative measure of dispersion.
Q.6 Differentiate between absolute and relative measures of dispersion.
Q.7 What are the different absolute measures of dispersion?
Q.8 Enlist the relative measures of dispersion.
Q.9 Write a short note on range.
Q.10 Define the range and how is it calculated for grouped data.
Q.11 State the advantages and disadvantages of the range.
Q.12 Given $X_m = 100$ and $X_0 = 25$. Find coefficient of range.
Ans. 0.6
Q.13 Given upper class boundary of the highest class = 80 and lower class boundary of the lowest class = 20. Compute coefficient of range.
Ans. 0.6
Q.14 Given mid value of the highest class = 75 and mid value of the lowest class = 15. Determine range and coefficient of range.
Ans. 60 and 0.67
Q.15 Define inter-quartile range.
Q.16 Write down the definition of quartile deviation.
Q.17 Define semi inter-quartile range and its coefficient.
Q.18 Given $Q_1 = 25$ and $Q_3 = 75$. Find quartile deviation.
Ans. 25
Q.19 Given $Q_1 = 125$ and $Q_3 = 175$. Find coefficient of quartile deviation.
Ans. 0.17
Q.20 In a symmetrical distribution $Q_1 = 140$ and median = 150. Find quartile deviation.
Ans. 10
Q.21 In a symmetrical distribution median = 50 and $Q_3 = 60$. Find quartile deviation.
Ans. 10
Q.22 Given quartile deviation = 11.25 and $Q_3 = 90.745$. Find coefficient of quartile deviation.
Ans. 0.14
Q.23 Given quartile deviation = 1.75 and $Q_1 = 40.25$. Find coefficient of quartile deviation.
Ans. 0.0417
Q.24 What is mean deviation and how is it calculated?
Q.25 Define average deviation and its coefficient.
Q.26 Given X = 2, 4, 6. Find mean deviation from median.
Ans. 4/3

Q.27 Given $X = 4, 6, 8, 8, 10$. Find mean deviation from mode.
Ans. 1.6

Q.28 Given $n = 6$, $\bar{X} = 4$ and $\sum|X - \bar{X}| = 24$. Find coefficient of mean deviation.
Ans. 1

Q.29 Given $n = 9$, mode $= 7$ and $\sum|X - \text{mode}| = 23$. Compute mode coefficient of dispersion.
Ans. 0.37

Q.30 Given median $= 8$, $n = 4$ and $\sum|X - \text{median}| = 48$. Compute coefficient of mean deviation from median.
Ans. 1.5

Q.31 The sum of absolute deviations from median 25 for 9 observations is 180. Find the coefficient of mean deviation.
Ans. 0.8

Q.32 Define the standard deviation.

Q.33 What is standard deviation and how is it calculated?

Q.34 Write down the various formulas of standard deviation.

Q.35 Write down the various formulas of variance.

Q.36 Define population variance.

Q.37 Explain the biased sample variance and unbiased sample variance.

Q.38 Write down the mathematical properties of standard deviation.

Q.39 Write down the mathematical properties of variance.

Q.40 Write short note on coefficient of variation.

Q.41 Define the coefficient of variation and coefficient of standard deviation.

Q.42 Find biased sample standard deviation of the scores 30, 35, 40.
Ans. 4.08

Q.43 Find unbiased sample standard deviation of the scores 30, 35, 40.
Ans. 5

Q.44 Given $X = 2, 4, 6, 8, 10$. Find biased sample variance.
Ans. 8

Q.45 Given $X = 2, 4, 6, 8, 10$. Find unbiased sample variance.
Ans. 10

Q.46 If $Var(X) = 25$, then find $Var(2X + 4)$ and $S.D.(4 - 2X)$
Ans. 100 and 10.

Q.47 Given $\Sigma X = 180$, $S^2 = 36$ and $n = 5$. Find ΣX^2.
Ans. 6660

Q.48 Given $\Sigma X = 180$, $\Sigma X^2 = 6660$ and $n = 5$. Find coefficient of variation.
Ans. 16.67%

Q.49 The mean and coefficient of variation of 20 observations are 20 and 25% respectively. Find variance.
Ans. 25

Q.50 The mean of 200 items is 48 and their standard deviation is 3. Find the sum of squares of all items.
Ans. 462600

Q.51 In a surprise checking of passengers in a local bus, 5 passengers without tickets were caught. The sum of squares and the standard deviation of the amount found in their pockets were Rs. 5000 and Rs. 10 respectively. If the total fine imposed is equal to the amount discovered from them, and fine imposed is uniform. Find amount of fine to be paid by each passenger.

Ans. Rs. 30

Q.52 Given $n = 100$, $\sum(X - 4) = -11$ and $\sum(X - 4)^2 = 257$. Compute coefficient of variation.

Ans. 41.13%

Q.53 Given $\bar{X} = 20$ and $Var(X) = 25$. If $Y = 2X - 40$ then find mean and variance of Y.

Ans. 0 and 100

Q.54 For a set of 10 observations, the sum of the deviations from 25 is 100 and the sum of the squares of these deviations is 1640. Find the coefficient of variation.

Ans. 22.86%

Q.55 The sum of squares of deviations taken from mean 50 for 10 sample observations is 640. Find the coefficient of variation.

Ans. 16%

Q.56 Write down the various measures of dispersion along with their formulae.

Q.57 Explain the moments about mean.

Q.58 Discuss the relation between the mean moments and the raw moments.

Q.59 Write first four raw moments in class interval units.

Q.60 Write the formulas of first four moments about origin.

Q.61 Write down the formulas for moments about the arbitrary origin.

Q.62 What is meant by moment ratios?

Q.63 Define the moment ratios (β_1 and β_2). For what purposes these are computed?

Q.64 The first two moments of a distribution about the value 5 of a variable are 2 and 32. Find variance.

Ans. 28

Q.65 The first three moments of a distribution about the value $X = 20$ are 1, 4, 10. Find the value of b_1.

Ans. $b_1 = 0$

Q.66 The first four moments about the arithmetic mean of a distribution are 0, 4, 6 and 48. Find b_2.

Ans. $b_2 = 3$

Q.67 The first two moments of a distribution about $X = 10$ are 2 and 20. Find coefficient of variation.

Ans. 33.33%

Q.68 The first moment about $X = 25$ is 5. Find arithmetic mean.

Ans. 30

Q.69 Given mean = 25, variance = 16 and $\beta_1 = 1$. Find the third moment about mean.

Ans. 64

Q.70 Write the formulas of corrected central moments.

Q.71 Explain the term skewness.

Q.72 Distinguish between positive and negative skewness with diagrams.

Q.73 What is meant by skewness and how it can be calculated?

Q.74 Distinguish between symmetry and skewness.

Q.75 Write down the various formulas for the measurement of skewness.

Q.76 Can the value of mean, median and mode be the same? If yes, state the situation.

Q.77 Can the difference between Q_3 and median be equal to the difference between median and Q_1? If yes, state the situation.

Q.78 Given mean = 100, mode = 95 and standard deviation = 10. Find coefficient of skewness.

Ans. 0.5

Q.79 Given mean = 50, median = 48 and standard deviation = 6. Find Karl Pearson's coefficient of skewness.

Ans. 1

Q.80 For a moderately skewed distribution the mean price is Rs.20, the median price is Rs.17 and the coefficient of variation is 20%. Find coefficient of skewness.

Ans. 2.25

Q.81 Given mean = 50, median = 48 and coefficient of skewness = 1. Find the value of the variance.

Ans. 36

Q.82 In a symmetrical distribution $Q_1 = 40$ and $Q_3 = 60$. Find median.

Ans. 50

Q.83 What would be the shape of the distribution if:

(a) mean = median = mode (b) mean > median > mode

(c) mean < median < mode (d) (Q_3 – median) > (median – Q_1)

Ans. (a) symmetrical (b) positively skewed

(c) negatively skewed (d) positively skewed

Q.84 Given Q_3 – median = 25 and median – Q_1 = 16. Compute Bowley's coefficient of skewness.

Ans. 0.22

Q.85 Define kurtosis.

Q.86 Define leptokurtic distribution.

Q.87 Define mesokurtic distribution.

Q.88 Define platykurtic distribution.

Q.89 Explain skewness and kurtosis.

EXERCISES

Q.1 Compute the range and the coefficient of range for the observations:

(i) 2000, 35, 400, 15, 40, 1500, 300, 6, 90, 250

(ii) 8, 11, 7, 15, 9, 12, 10, 6

(iii) 2, 3, 3, 5, 5, 5, 8, 10, 12

(iv) $X_1 = 1, X_2 = 10, X_3 = 2, X_4 = 9, X_5 = 3, X_6 = 8, X_7 = 4, X_8 = 7, X_9 = 5, X_{10} = 6$.

Ans. (i) 1994, 0.994 (ii) 9, 0.429 (iii) 10, 0.714 (iv) 9, 0.818.

Q.2 Find the range and coefficient of the range of the following data:

Values of variable	2	4	6	8	10
Frequency	3	10	25	45	27

Ans. 8, 0.667

Q.3 Find the range and coefficient of range from the following frequency distribution.

Weights (grams)	65 – 84	85 – 104	105 – 124	125 – 144	145 – 164	165 – 184
No. of apples	9	10	17	10	8	6

Ans. 120, 0.482

Q.4 For the data given below, compute quartile deviation and coefficient of quartile deviation: 1950, 1870, 1870, 1775, 1745, 1720, 1740, 1670, 1710, 1590, 1030, 1110, 1070, 1190, 1230, 1310, 1350, 1332, 1430, 1460.

Ans. 246.875, 0.165

Q.5 Calculate the quartile deviation and its coefficient from the following data:

Roll No.	1	2	3	4	5	6	7	8	9	10	11	12
Marks	39	41	41	40	40	43	43	42	42	44	44	45

Ans. 1.75, 0.042.

Q.6 Consider the following data:

Hourly wages (Rs.)	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
Number of persons	2	5	6	5	2

Compute the quartile deviation and coefficient of quartile deviation.

Ans: 9, 0.14

Q.7 Calculate the quartile deviation and coefficient of quartile deviation from the distribution of aptitude scores.

Scores	50 – 59	60 – 69	70 – 79	80 – 89	90 – 99
No. of applicants	7	12	15	4	2

Ans: 7.415, 0.11

Q.8 Compute the semi-inter-quartile range for the cholesterol level of the 20 patients at a hospital.

Cholesterol level	200 – 204	205 – 209	210 – 214	215 – 219	220 – 224
Number of patients	3	4	7	5	1

Ans: 4.25

Q.9 Calculate the mean deviation from mean and mean deviation from median for the following data:2, 6, 9, 12, 8, 13, 5, 6, 23, 16.

Ans. 4.8, 4.6

Q.10 Calculate the mean deviation from (i) arithmetic mean, (ii) median, (iii) mode in respect of the marks obtained by nine students given below and show that the mean deviation from median is minimum. Marks (out of 25): 7, 4, 10, 9, 15, 12, 7, 9, 7. If the marks are doubled (converted out of 50), will the variation in marks increase? Give reasons.

Ans. (i) 2.35 (ii) 2.33 (iii) 2.56.

Q.11 Calculate the mean deviation about mean from the following distribution of differences in ages between husbands and wives in a particular community:

Difference (in years)	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25
Frequency	140	200	100	50	10

Ans: 4.024

Q.12 For the following frequency table, calculate the mean deviation about median and its coefficient.

Height in inches	55	56	57	58	59	60
Number of boys	25	40	50	20	10	5

Ans: 1.01, 0.02

Q.13 Following are the monthly salaries of a sample of 30 workers in an office. Compute the mean deviation from the mode. Also compute the mode coefficient of dispersion.

Monthly salary (in thousand Rs.)	Number of workers
6 and under 8	6
8 and under 10	5
10 and under 12	10
12 and under 14	6
14 and under 16	3

Ans: 1.98, 0.18

Q.14 Calculate mean deviation with mean and median of the marks of the 80 students.

Marks	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
No. of students	3	8	9	15	20	13	8	4

Ans. 14.188, 14.063

Q.15 Ten measurements gave the following results:

Length in cms: 77, 73, 75, 70, 72, 76, 75, 72, 74, 76.

Find variance and standard deviation.

Ans. 4.4, 2.098.

Q.16 Find standard deviation and coefficient of variation of the numbers: 3.2, 4.6, 2.8, 5.2, 4.4.

Ans. 0.90, 22.28 %

Q.17 Calculate the variance of 14, 16, 18, 20, 22. Without recalculation, find the variance after (i) adding 5 to each observation (ii) subtracting 5 from each observation (iii) multiplying each observation by 2 (iv) dividing each observation by 2 (v) multiplying each observation by 2 and adding 5.

Ans. 8 (i) 8 (ii) 8 (iii) 32 (iv) 2 (v) 32

Q.18 Compute the coefficient of variation of (i) X (ii) Y = 2X where X has values, 2, 3, 3, 5, 5, 5, 8, 10, 12. Are the two results same? If so, give reason.

Ans. (i) 54.5 % (ii) 54.5 %.

Reason: Coefficient of variation is independent of change of scale.

Q.19 Find the arithmetic mean and the standard deviation of the observations: 40, 40, 50, 60, 70, 70, 80, 80, 90. Also find the arithmetic mean and the standard deviation after increasing the observations by (i) 10 units (ii) 10 percent.

Ans. 64.44, 17.07 (i) 74.44, 17.07 (ii) 70.89, 18.78

Q.20 What will be the standard deviation and the variance in each of the following cases? (i) 2X (ii) X + 2 (iii) 2X + 4. If Var (X) = 25.

Ans. (i) 10, 100 (ii) 5, 25 (iii) 10, 100.

Q.21 Given the following information:

$$n_1 = 150, \quad \Sigma(X_1 - 100) = 180, \quad \Sigma(X_1 - 100)^2 = 245320$$

$$n_2 = 200, \quad \Sigma(X_2 - 100) = 250, \quad \Sigma(X_2 - 100)^2 = 43850$$

Compare the variability by calculating the coefficient of variation of each series.

Ans. C.V.(1) = 39.94 %, C.V.(2) = 14.57 %.

Q.22 Rawat electronics is giving two training programmes? Two groups were trained for the same task. For the first group trained by programme A, it took a mean of 28.74 hours with a variance of 79.39. For the second group trained by programme B, it took a mean of 20.5 hours with a variance of 54.76. Which training programme has less relative variability in its performance?

Ans. Programme A

Q.23 The weight measurements in grams of 20 eggs are given below:

65, 68, 60, 76, 65, 74, 58, 56, 61, 57, 63, 64, 72, 66, 65, 67, 65, 63, 67, 67.

Find the percentage of observations lying within the limits

(i) $\bar{X} \pm S$ (ii) $\bar{X} \pm 2S$ (iii) $\bar{X} \pm 3S$.

Ans. (i) 70% (ii) 95% (iii) 100%

Q.24 A professor of statistics wants to estimate the average study time per week for students in statistics courses at his college. To accomplish this, he selects a sample of 25 students and records their weekly study times. The times, to the nearest hour, are:

Study time	3	4	5	6	7	8
Number of students	2	3	7	8	3	2

Compute the standard deviation.

Ans: 1.30

Q.25 A frequency distribution on the length of telephone calls monitored at the switchboard of an office is given below:

Length of calls (minutes)	Number of calls
0 and under 2	5
2 and under 4	10
4 and under 6	40
6 and under 8	30
8 and under 10	15

Compute the variance and coefficient of variation of the calling time.

Ans: 4.16, 35.17 %

Q.26 Calculate standard deviation taking :

(i) $D = X - 17.5$ and (ii) $u = \dfrac{X - 22.5}{5}$.

Group	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35
f	6	18	18	18	18	18

Ans. (i) 7.75 (ii) 7.75

Q.27 Compute the variance from the following data using step deviation method.

X	200	300	350	700	840
f	4	2	2	1	1

Ans: 45564

Q.28 The goals scored by two teams A and B in a football season were as follows:

No. of goals scored in a match		0	1	2	3	4
Number of matches	Team A	27	9	8	5	4
	Team B	17	9	6	5	4

Which team should be considered as more consistent.

Ans. Team B

Q.29 An analysis of the daily wages paid to workers in two firms A and B belonging to the same industry, gives the following results.

	Firm A	Firm B
No. of workers	160	150
Average wage	269	275
Variance of wage	100	121

Find out: (i) which firm pays larger amount as daily wages?

(ii) In which firm is there greater variability in individual wages?

Ans. (i) Firm A (ii) Firm B.

Q.30 You are provided with the following raw sums in a statistical investigation of two variables, X and Y:

$$\Sigma X = 235, \quad \Sigma Y = 250, \quad \Sigma X^2 = 6750, \quad \Sigma Y^2 = 6840$$

Ten pairs of values are included in the survey. Compute the standard deviation of the X and Y variables.

Ans. S.D.(X) = 11.08, S.D.(Y) = 7.68

Q.31 The mean of 200 items is 48 and their standard deviation is 3. Find the sum and sum of squares of all these items.

Ans. $\Sigma X = 9600$, $\Sigma X^2 = 462600$

Q.32 A computer calculated mean and standard deviation from 20 observations as 42 and 5 respectively. It was later discovered at the time of checking that it had copied down two values as 45 and 38, whereas the correct values were 35 and 58. Find correct value of coefficient of variation.

Ans. 14.65 %

Q.33 A student calculated the values of mean and standard deviation of 25 observations as 20 and 4 respectively. It was later discovered at the time of checking that he had copied two values as 7 and 18 while the correct values were 13 and 17. Find the correct value of coefficient of variation.

Ans. 16.68 %

Q.34 Mean and variance of marks of section A are 62 and 16 respectively and that of section B are 69 and 25 respectively while the number of students are 30 and 40. Find out combined mean. Which section has greater absolute dispersion? Which has greater relative dispersion?

Ans. 66, B has greater absolute dispersion, B has greater relative dispersion.

Q.35 Given the following data. Determine the mean and standard deviation of the combined set:

No. of items	Mean	Variance
25	25.2	4.90
30	21.5	6.25

Ans. 23.18, 3

Q.36 The following are some of the particulars of the distribution of weights of boys and girls in a class.

	Boys	Girls
Number	100	50
Mean weight	60 kg	45 kg
Variance	9 kg^2	4 kg^2

Find arithmetic mean and standard deviation of the boys and girls taken together.

Ans. 55, 7.572

Q.37 A calculating machine while calculating mean and standard deviation of 25 readings misread one observation as 36 instead of 26. The following results were given by the machine. S.D. = 5 and $\bar{X} = 30$. What are the correct values of $\bar{X}$ and S.D.?

Ans. 29.6, 4.9

Q.38 A group has arithmetic mean of 10 from 60 items with variance 4. A subgroup of 40 items has mean 11 and standard deviation 1.5. Find the mean and standard deviation of other subgroup.

Ans. 8, 1.225

Q.39 The mean of 5 observations is 4.8 and variance is 4.56. If three of the five observations are 2, 5 and 6, find the other two observations.

Ans. 8 and 3.

Q.40 Calculate the first four moments about the mean for the observations

81, 87, 90, 93, 94, 95, 98, 104, 105.

Ans. 0, 52.54, – 40.56, 6249.26

Q.41 Calculate the first four moments about the mean for the following data. Also calculate b_1 and b_2 .

X	12	13	14	15	16	17	18	20
f	4	11	32	21	15	8	5	4

Ans. 0, 3.04, 4.86, 35.80; 0.84, 3.87.

Q.42 Calculate first four moments about X = 0 from the following data. Also calculate the moments about the mean.

Groups	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25
Frequency	1	3	5	2	1

Ans. 12.08, 172.92, 2763.02 , 47955.73; 0, 26.99, 21.98, 1964.62

Q.43 Given the following information:

$\Sigma f = 76$, $\Sigma fX = 572$, $\Sigma fX^2 = 4848$, $\Sigma fX^3 = 44240$, $\Sigma fX^4 = 425280$.

Calculate first four moments about the mean and test the distribution for symmetry and normality.

Ans. 0, 7.09, – 4.99, 119.36; $b_1 = 0.07$, $b_2 = 2.37$

Q.44 Find the first four moments about the mean for the data given below taking $u = \dfrac{X - 22}{5}$.

Groups	10 – 14	15 – 19	20 – 24	25 – 29	30 – 34	35 – 39
Frequency	2	4	6	8	3	1

Ans. 0, 39.25, – 26.25, 3843.75

Q.45 The first three moments of a distribution about the value 2 of a variable are 1, 16 and – 40. Show that the mean is 3, variance is 15 and $m_3 = -86$.

Q.46 The first three moments of a distribution about the value X = 4 are 1, 4 and 10. Is the distribution symmetrical, positively skewed or negatively skewed?

Ans. Yes, the distribution is symmetrical.

Q.47 The first four moments of a distribution about X = 2 are 1, 2.5, 5.5 and 16. Calculate the mean and the coefficient of variation.

Ans. $\bar{X} = 3$, C. V. = 40.67 %

Q.48 The first four moments of a set of numbers about 3 are – 2, 10, – 25 and 50. Determine the corresponding moments about the mean.

Ans. $m_1 = 0$, $m_2 = 6$, $m_3 = 19$, $m_4 = 42$

Q.49 The first four moments about X = 17.5 of a distribution are 0.3, 74, 45 and 12125. Find out whether the distribution is leptokurtic or platykurtic.

Ans. platykurtic.

Q.50 Following are the data relating to weights of students. Calculate the coefficient of skewness by Karl Pearson's method.

Weights (pounds)	118 – 126	127 – 135	136 – 144	145 – 153	154 – 162
Number of students	2	6	10	4	3

Ans: 0.09

Q.51 Compute the coefficient of skewness using the averages and standard deviation.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	4	9	12	8	2

Ans. – 0.05

Q.52 Find the Bowley's coefficient of skewness for the data given below:

Groups	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69
Frequency	14	25	40	32	17	10

Ans. 0.044

Q.53 For each of the following unimodal distributions examine whether the distribution is symmetrical, skewed to the left or skewed to the right.

(i) Mean = 19, Mode = 52

(ii) Q_1 = 136, Median = 160, Q_3 = 184

(iii) Mean = 78, Median = 61

(iv) Q_1 = 14, Median = 29, Q_3 = 53.

Ans. (i) left skewed (ii) symmetrical (iii) right skewed (iv) right skewed

Q.54 From the marks secured by 120 students in section A and 120 students of section B of a class, the following measures are obtained.

Section A	A.M. = 46.83	Mode = 51.67	S = 14.8
Section B	A.M. = 47.83	Mode = 47.07	S = 14.8

Describe which distribution of marks is more skewed.

Ans. Coefficient of skewness (section A) = – 0.327.
Coefficient of skewness (section B) = 0.051

Distribution A is more skewed than distribution B.

Q.55 (i) Given: $\Sigma f = 120$, $\Sigma fX = 296$, Mode = 2.944 and second moment about mean is 1.4884. Calculate coefficient of skewness.

(ii) The lower and upper quartiles of a distribution are 142.36 and 167.73 respectively, while median is 153.50. Find coefficient of skewness.

Ans. (i) – 0.39 (ii) 0.122

Chapter

5

INDEX NUMBERS

5.1. INTRODUCTION

Index numbers are meant to study the changes in the effects of such factors which cannot be measured directly. According to Bowley, "Index numbers are used to measure the changes in some quantity which we cannot observe directly". For example, changes in business activity in a country are not capable of direct measurement but it is possible to study relative changes in business activity by studying the variations in the values of some such factors which affect business activity, and which are not capable of direct measurement.

Index numbers are commonly used statistical devices for measuring the combined fluctuations in a group of related variables. If we wish to compare the price level of consumer items today with that prevalent ten years ago, we are not interested in comparing the prices of only one item, but in comparing some sort of average price levels. We may wish to compare the present agricultural production or industrial production with that at the time of independence. Here again, we have to consider all items of production and each item may have undergone a different fractional increase (or even a decrease). How do we obtain a composite measure? This composite measure is provided by index numbers which may be defined as a device for combining the variations that have come in a group of related variables over a period of time, with a view to obtain a figure that represents the 'net' result of the change in the constituent variables.

Index numbers may be classified in terms of the variables that they are intended to measure. In business, different groups of variables in the measurement of which index number techniques are commonly used are (i) price (ii) quantity (iii) value and (iv) business activity. Thus, we have index of wholesale prices, index of consumer prices, index of industrial output, index of value of exports and index of business activity etc. Here we shall be mainly interested in index numbers of prices showing changes with respect to time, although methods described can be applied to other cases. In general, the present level of prices is compared with the level of prices in the past. The present period is called the current period and some period in the past is called the base period.

5.2. INDEX NUMBERS

Index numbers are statistical measures designed to show changes in a variable or group of related variables with respect to time, geographic location or other characteristics such as income, profession, etc. A collection of index numbers for different years, locations, etc., is sometimes called an index series.

5.3. SIMPLE INDEX NUMBER

A simple index number is a number that measures a relative change in a single variable with respect to a base.

5.4. COMPOSITE INDEX NUMBER

A Composite index number is a number that measures an average relative change in a group of related variables with respect to a base.

5.5. TYPES OF INDEX NUMBERS

Following types of index numbers are usually used:

5.5.1. Price Index Numbers

Price index numbers measure the relative changes in prices of a commodities between two periods. Prices can be either retail or wholesale.

5.5.2. Quantity Index Numbers

These index numbers are constructed to measure changes in the physical quantity of goods produced, consumed or sold of an item or a group of items.

5.6. USES OF INDEX NUMBERS

The main uses of index numbers are given below:

(i) Index numbers are used in the fields of commerce, , meteorology, labour, industries etc.

(ii) The index numbers measure fluctuations during intervals of time, group differences of geographical position or degree etc.

(iii) They are used to compare the total variations in the prices of different commodities in which the unit of measurements differ with time and place etc.

(iv) They measure the purchasing power of money.

(v) They are helpful in forecasting the future economic trends.

(vi) They are used in studying difference between the comparable categories of animals, persons or items.

(vii) Index numbers of industrial production are used to measure the changes in the level of industrial production in the country.

(viii) Index numbers of import prices and export prices are used to measure the changes in the trade of a country.

(ix) The index numbers are used to measure seasonal variations and cyclical variations in a time series.

5.7. LIMITATIONS OF INDEX NUMBERS

(i) They are simply rough indications of the relative changes.

(ii) The choice of representative commodities may lead to fallacious conclusions as they are based on samples.

(iii) There may be errors in the choice of base periods or weights etc.

(iv) Comparisons of changes in variables over long periods are not reliable.

(v) They may be useful for one purpose but not for the other.

(vi) They are specialized types of averages and hence are subject to all those limitations with which an average suffers from.

5.8. MAIN STEPS TO CONSTRUCT PRICE INDEX NUMBERS

The following steps are considered for the construction of price index numbers:

5.8.1. Object

The first and the most important step in the construction of index numbers is to decide the object of making the index numbers of prices. The prices may be retail or whole-sale. The index numbers of retail prices are called the consumer price index (CPI) numbers and if the whole-sale prices are taken into consideration, the index numbers are called the whole-sale price index numbers. The index numbers of prices may be calculated for a certain locality, for a certain class of people like textile workers or office clerks etc. The index numbers may be required for geographical regions like districts or provinces etc. First of all we decide the purpose of making the index numbers. Once the purpose is decided, then we decide about the scope and the area or the people who are to be considered.

5.8.2. Selection of Commodities

A list of important commodities is prepared. Those commodities are taken into account which are commonly consumed by the consumers. There is no hard and fast rule about the number of commodities. Only those commodities are considered on which a reasonable amount is spent. The commodities on which the expenditure is meager or they are used only occasionally are not included in the list. Thus the commodities which are representative of the tastes and customs of the people are taken in the list. Dr. Irving Fisher has said that 20 commodities is a small number and 50 commodities is a reasonable number. For construction of wholesale price index numbers, about 80 commodities are taken in the list and for retail-price index numbers about 300 commodities are considered. Sometimes the index numbers of very important commodities like wheat, rice, oil, ghee etc. are calculated. These index numbers are based on about one dozen commodities and are called sensitive price index numbers.

5.8.3. Collection of Price Data

The most important and difficult step is the collection of prices. The prices are to be taken from the field. For retail price index numbers, retail prices are needed. The prices change from place to place and from time to time. On different shops the prices are different. In actual practice there are many difficulties. Usually some representative shops from where the consumers mostly purchase their items are selected and the prices are taken from those shops. The prices are taken on daily basis and then the weekly and monthly averages are calculated. Finally quarterly or

yearly averages are calculated. Some commodities are sold in different varieties. Rice, sugar, mangoes etc. have different varieties which are sold on different prices. This problem is solved by assigning due weights to different varieties and then weighted average price is calculated. Sometimes different varieties are treated as different commodities.

For whole-sale prices, the prices are taken from the whole-sale markets, factories, depots and the whole-sale agencies. The whole-sale prices are usually stable, therefore these prices are not taken on daily basis. The price reporting is done on weekly or monthly basis depending upon the nature of the commodity. The prices of some commodities are controlled by the government. These prices are reported whenever some change takes place.

5.8.4. Selection of Base Period

The prices of the commodities in the current period are to be compared with the prices of some period in the past. This period in the past is called the base period or the reference period. The base period is decided by the experts. For the construction of national index numbers, the base period is decided by statistics division, Government of Pakistan. This period should not be in the remote past. The period which is economically stable and is free of disturbances and strikes is taken as the base period.

There are two methods of selecting the base period. These are fixed base method and chain base method. They are discussed below:

(a) Fixed Base Method

In fixed base method, a particular year is generally chosen arbitrarily and the prices of the subsequent years are expressed as relatives of the prices of the base year. Sometimes instead of choosing a single year as the base, a period of a few years is chosen and the average price of this period is taken as the base year's price. The year which is selected as a base should be a normal year or in other words, the price level in this year should neither be abnormally low nor abnormally high. If an abnormal year is chosen as the base, the price relatives of the current year calculated on its basis would give misleading conclusions. For example, a year in which war was at its peak, say the year 1965, is chosen as a base year, the comparison of the price level of subsequent years to the prices of 1965 is bound to give misleading conclusions. The reason is that the price level in the year 1965 was abnormally high. In order to remove this difficulty associated with the selection of a normal year, the average price of a few years is sometimes taken as the base price. The fixed base method is used by the Government in the calculation of national index numbers.

In Fixed Base,

$$\text{Price relative for current year} = \frac{\text{Price of current year}}{\text{Price of base year}} \times 100$$

$$\text{or} \quad P_{on} = \frac{p_n}{p_o} \times 100$$

Example 5.1.

Given the prices of rice (Basmati) in Multan in rupees per quintal from 1987 to 1992. Calculate the simple index numbers with 1987 as base.

Years	1987	1988	1989	1990	1991	1992
Prices of rice (Rs. per quintal)	618.25	741.00	749.00	789.00	927.00	1034.00

Source: Directorate of Agriculture (Economics and Marketing) Punjab, Lahore.

Solution:

The necessary calculations are given below:

Years	Prices (Rs.) p_n	Index No. = $\frac{p_n}{p_o} \times 100$ (1987 as base)
1987	618.25	100
1988	741.00	$\frac{741.00}{618.25} \times 100 = 119.85$
1989	749.00	$\frac{749.00}{618.25} \times 100 = 121.15$
1990	789.00	$\frac{789.00}{618.25} \times 100 = 127.62$
1991	927.00	$\frac{927.00}{618.25} \times 100 = 149.94$
1992	1034.00	$\frac{1034.00}{618.25} \times 100 = 167.25$

Example 5.2.

The following are average whole-sale prices of wheat (Maxi-Pak) in Lahore in rupees per quintal for the period 1983 to 1992. Compute the index numbers using (i) average of first four years as base (ii) average of all the ten years as base.

Years	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Prices (Rs.)	186.25	195.00	200.00	210.19	230.31	237.50	243.00	287.00	294.00	363.00

Source: Bureau of Statistics Government of the Punjab Lahore.

Solution:

The necessary calculations are given below:

Years	Prices (Rs.) p_n	(i) Price relatives $= \frac{p_n}{p_o} \times 100$ (Average of first 4 years as base)	(ii) Price relatives $= \frac{p_n}{p_o} \times 100$ (Average of 10 years as base)
1983	186.25	$\frac{186.25}{197.86} \times 100 = 94.13$	$\frac{186.25}{244.63} \times 100 = 76.14$
1984	195.00	$\frac{195.00}{197.86} \times 100 = 98.55$	$\frac{195.00}{244.63} \times 100 = 79.71$
1985	200.00	$\frac{200.00}{197.86} \times 100 = 101.08$	$\frac{200.00}{244.63} \times 100 = 81.76$
1986	210.19	$\frac{210.19}{197.86} \times 100 = 106.23$	$\frac{210.19}{244.63} \times 100 = 85.92$
1987	230.31	$\frac{230.31}{197.86} \times 100 = 116.40$	$\frac{230.31}{244.63} \times 100 = 94.15$
1988	237.50	$\frac{237.50}{197.86} \times 100 = 120.03$	$\frac{237.50}{244.63} \times 100 = 97.09$
1989	243.00	$\frac{243.00}{197.86} \times 100 = 122.81$	$\frac{243.00}{244.63} \times 100 = 99.33$
1990	287.00	$\frac{287.00}{197.86} \times 100 = 145.05$	$\frac{287.00}{244.63} \times 100 = 117.32$
1991	294.00	$\frac{294.00}{197.86} \times 100 = 148.59$	$\frac{294.00}{244.63} \times 100 = 120.18$
1992	363.00	$\frac{363.00}{197.86} \times 100 = 183.46$	$\frac{363.00}{244.63} \times 100 = 148.39$
	$\Sigma p_n = 2446.25$		

(i) The average of first four years is

$$\frac{186.25 + 195.00 + 200.00 + 210.19}{4} = \frac{791.44}{4} = 197.86 = p_o$$

(ii) The average of all the ten years is

$$\frac{\Sigma p_n}{n} = \frac{2446.25}{10} = 244.625 \text{ or } 244.63 = p_o$$

(b) Chain Base Method

In this method, there is no fixed base period. The year immediately preceding the one for which price index have to be calculated is assumed as the base year. Thus, for the year 1994 the base year would be 1993, for 1993 it would be 1992, for 1992 it would be 1991 and so on. In this way there is no fixed base. It goes on changing. The chief advantage of this method is that the price relatives of a year can be compared with the price level of the immediately preceding year. Businessmen are mostly interested in comparison of this type rather than in comparisons relating to distant past. Yet another advantage of the chain base method is that it is possible to include new items in an index number or to delete old items which are no more important. In fixed base method it is not possible. But chain base method has a drawback that comparisons cannot be made over a long period.

In Chain Base,

$$\text{Link relative of current year} = \frac{\text{Price in the current year}}{\text{Price in the preceding year}} \times 100$$

or in symbols, $p_{n-1,n} = \dfrac{p_n}{p_{n-1}} \times 100$

Example 5.3.

Compute chain indices, taking 1987 as base from the data given in Example 5.1.

Solution:

The necessary calculations are given below:

Years	Prices of rice (Rs. per quintal)	Link relatives $\frac{p_n}{p_{n-1}} \times 100$	Chain indices
1987	618.25	100	100
1988	741.00	$\frac{741.00}{618.25} \times 100 = 119.85$	$\frac{100 \times 119.85}{100} = 119.85$
1989	749.00	$\frac{749.00}{741.00} \times 100 = 101.08$	$\frac{119.85 \times 101.08}{100} = 121.14$
1990	789.00	$\frac{789.00}{749.00} \times 100 = 105.34$	$\frac{121.14 \times 105.34}{100} = 127.61$
1991	927.00	$\frac{927.00}{789.00} \times 100 = 117.49$	$\frac{127.61 \times 117.49}{100} = 149.93$
1992	1034.00	$\frac{1034.00}{927.00} \times 100 = 111.54$	$\frac{149.93 \times 111.54}{100} = 167.23$

5.8.5. Selection of the Suitable Average

There are different averages which can be used in averaging the price relatives or link relatives of different commodities. Experts have suggested that the geometric mean should be calculated for averaging these relatives. But as the calculation of the geometric mean is difficult, it is mostly avoided and the arithmetic mean is commonly used. In some cases the median is used to remove the effect of the wild observations.

5.8.6. Selection of Suitable Weights

In calculation of price index numbers all commodities are not of equal importance. In order to give them their due importance, commodities are given due weights. Weights are of two kinds (a) Implicit weights, (b) Explicit weights. In the first kind the weights are not explicitly assigned to any commodity but the commodity to which greater importance is attached is repeated a number of times. A number of varieties of such commodities are included in the index number as separate items. Thus, if in an index number wheat is to receive a weight of 3 and rice a weight of 2, three varieties of wheat and two varieties of rice would be included. In this method weights are not apparent, but items are implicitly weighted. Such weights are known as implicit weights. In the second kind, weights are explicitly assigned to commodities. Only one variety of a commodity is included in the construction of index number but its price relative is multiplied by the figure of weight assigned to it. Explicit weights are decided on some logical basis. For example, if wheat and rice are to be weighted in accordance with the value of their net output and if the ratio of their net output is 5 : 2, wheat would receive a weight of five and rice two. Such weights are called explicit weights. Sometimes the quantities which are consumed are used as weights. These are called quantity weights. The amount spent on different commodities can also be used as their weights. These are called the value weights.

5.9. UNWEIGHTED INDEX NUMBERS

There are two methods of constructing unweighted index numbers:

5.9.1. Simple Aggregative Method

It is obtained by dividing the sum of the prices of current year by the sum of the prices of base year and expressed as percentage i.e.,

$$P_{on} = \frac{\Sigma p_n}{\Sigma p_o} \times 100$$

Where, Σp_n represents the sum of prices for the current year, and Σp_o represents the sum of prices for the base year.

5.9.2. Simple Average of Relatives

In simple average of relatives we find price relatives of the given commodities and then average them. Any one of the average, mean, median and geometric mean may be used. If we use arithmetic mean, we have

$$P_{on} = \frac{1}{N}\Sigma\left(\frac{p_n}{p_o}\right) \times 100$$

Where, N represents the number of commodities.

Example 5.4.

The following are the prices of four different commodities for 1990 and 1991. Compute a price index by (i) Simple aggregative method and (ii) Average of price relatives method by using both arithmetic mean and geometric mean, taking 1990 as base.

Commodity	Kappas (American)	Wheat (Maxi-Pak)	Rice (Basmati)	Gram (Whole)
Price in 1990 (Rs. per quintal)	909	288	767	659
Price in 1991 (Rs. per quintal)	874	305	910	573

Source: Bureau of Statistics Government of the Punjab Lahore.

Solution:

The necessary calculations are given below:

Commodity	Price in 1990 (Rs. per quintal) p_o	Price in 1991 (Rs. per quintal) p_n	Price relatives $P = \frac{p_n}{p_o} \times 100$	log P
Kappas	909	874	$\frac{874}{909} \times 100 = 96.15$	1.9829
Wheat	288	305	$\frac{305}{288} \times 100 = 105.90$	2.0249
Rice	767	910	$\frac{910}{767} \times 100 = 118.64$	2.0742
Gram	659	573	$\frac{573}{659} \times 100 = 86.95$	1.9393
Total	$\Sigma p_o = 2623$	$\Sigma p_n = 2662$	$\Sigma\left(\frac{p_n}{p_o}\right) \times 100 = 407.64$	$\Sigma \log P = 8.0213$

(i) Simple Aggregative Method:

$$P_{on} = \frac{\Sigma p_n}{\Sigma p_o} \times 100 = \frac{2662}{2623} \times 100 = 101.49$$

(ii) Average of Price Relatives Method (using arithmetic mean):

$$P_{on} = \frac{1}{N}\Sigma\left(\frac{p_n}{p_o}\right) \times 100 = \frac{1}{4}(407.64) = 101.91$$

Average of Price Relatives Method (using geometric mean)

$$P_{on} = \text{Antilog}\left(\frac{\Sigma \log P}{N}\right) = \text{Antilog}\left(\frac{8.0213}{4}\right) = 101.23$$

Example 5.5.

The following table gives the average whole-sale prices in rupees per quintal of Kappas (American), Wheat (Maxi-Pak) and Rice (Basmati), during the year 1987 to 1991. Construct index numbers with 1987 as base using: (i) Arithmetic Mean (ii) Median as an average.

Year	Commodities (Prices in rupees per quintal)		
	Kappas (American)	Wheat (Maxi-Pak)	Rice (Basmati)
1987	620.00	204.06	616.25
1988	618.00	219.00	671.00
1989	719.00	245.00	704.00
1990	909.00	288.00	767.00
1991	874.00	305.00	910.00

Source:- Bureau of Statistics Government of the Punjab Lahore.

Solution:

The necessary calculations are given below:

Year	Price Relatives			Index Numbers (1987 as base)		
	Kappas (American)	Wheat (Maxi-Pak)	Rice (Basmati)	Total	(i) Mean	(ii) Median
1987	100	100	100	300	100	100
1988	$\frac{618.00}{620.00} \times 100 = 99.68$	$\frac{219.00}{204.06} \times 100 = 107.32$	$\frac{671.00}{616.25} \times 100 = 108.88$	315.88	105.29	107.32
1989	$\frac{719.00}{620.00} \times 100 = 115.97$	$\frac{245.00}{204.06} \times 100 = 120.06$	$\frac{704.00}{616.25} \times 100 = 114.24$	350.27	116.76	115.97
1990	$\frac{909.00}{620.00} \times 100 = 146.61$	$\frac{288.00}{204.06} \times 100 = 141.13$	$\frac{767.00}{616.25} \times 100 = 124.46$	412.20	137.40	141.13
1991	$\frac{874.00}{620.00} \times 100 = 140.97$	$\frac{305.00}{204.06} \times 100 = 149.47$	$\frac{910.00}{616.25} \times 100 = 147.67$	438.11	146.04	147.67

Example 5.6.

The following table gives the prices of two commodities, Gold Tezabi and Silver Tezabi, during the year 1987 – 88 to 1991 – 92. Compute chain index numbers using 1987 – 88 price as base.

	Commodities Prices in Rupees	
Year	Gold Tezabi (per 10 grams)	Silver Tezabi (per kilogram)
1987 – 88	3045	4970
1988 – 89	2892	5333
1989 – 90	2972	5193
1990 – 91	3080	4758
1991 – 92	3140	4342

Source: Federal Bureau of Statistics

Solution:

The necessary calculations are given below:

	Link relatives				
Year	Gold Tezabi	Silver Tezabi	Total	Mean	Chain Indices
1987 – 88	100	100	200	100	100
1988 – 89	$\frac{2892}{3045} \times 100$ $= 94.98$	$\frac{5333}{4970} \times 100$ $= 107.30$	202.28	101.14	$\frac{100 \times 101.14}{100}$ $= 101.14$
1989 – 90	$\frac{2972}{2892} \times 100$ $= 102.77$	$\frac{5193}{5333} \times 100$ $= 97.37$	200.14	100.07	$\frac{101.14 \times 100.07}{100}$ $= 101.21$
1990 – 91	$\frac{3080}{2972} \times 100$ $= 103.63$	$\frac{4758}{5193} \times 100$ $= 91.62$	195.25	97.63	$\frac{101.21 \times 97.63}{100}$ $= 98.81$
1991 – 92	$\frac{3140}{3080} \times 100$ $= 101.95$	$\frac{4342}{4758} \times 100$ $= 91.26$	193.21	96.61	$\frac{98.81 \times 96.61}{100}$ $= 95.46$

Example 5.7.

Find chain indices from the following price relatives of three commodities, using geometric mean as average.

Year	Commodities A	B	C
1999	255	216	330
2000	186	162	384
2001	312	261	333
2002	279	225	462
2003	180	129	495

Solution:

The necessary calculations are given below:

Year	Link relatives A	B	C	G.M. $= [X_1.X_2.X_3]^{1/3}$	Chain Indices
1999	255	216	330	262.93	262.93
2000	$\frac{186}{255} \times 100 = 72.94$	$\frac{162}{216} \times 100 = 75.00$	$\frac{384}{330} \times 100 = 116.36$	86.02	$\frac{262.93 \times 86.02}{100} = 226.17$
2001	$\frac{312}{186} \times 100 = 167.74$	$\frac{261}{162} \times 100 = 161.11$	$\frac{333}{384} \times 100 = 86.72$	132.83	$\frac{226.17 \times 132.83}{100} = 300.42$
2002	$\frac{279}{312} \times 100 = 89.42$	$\frac{225}{261} \times 100 = 86.21$	$\frac{462}{333} \times 100 = 138.74$	102.27	$\frac{300.42 \times 102.27}{100} = 307.24$
2003	$\frac{180}{279} \times 100 = 64.52$	$\frac{129}{225} \times 100 = 57.33$	$\frac{495}{462} \times 100 = 107.14$	73.45	$\frac{307.24 \times 73.45}{100} = 225.67$

5.10. WEIGHTED INDEX NUMBERS

There are two methods of calculating weighted index numbers.

5.10.1. Weighted Aggregative Price Index Numbers

In this method we calculate the total expenditure of the current year and the base year. Price of each commodity is multiplied with the weight of the commodity which is usually the quantity consumed or quantity produced. The quantity of the

base year (q_o) or current year (q_n) can be used as weight. The current period expenditure is compared with the base period expenditure. The formula used is

$$\text{Price Index} = \frac{\text{Current year expenditure}}{\text{Base year expenditure}} \times 100$$

There are various kinds of weighted aggregative index numbers, some important formulas are discussed below:

(i) **Laspeyre's Index:** $P_{on} = \dfrac{\Sigma p_n\, q_o}{\Sigma p_o\, q_o} \times 100$

This is also called base year weighting system because the base year quantities q_o are used as weights.

(ii) **Paasche's Index:** $P_{on} = \dfrac{\Sigma p_n\, q_n}{\Sigma p_o\, q_n} \times 100$

This is also called current year weighting system because current year quantities q_n are used as weights.

(iii) Fisher's Ideal Index:

It is the geometric mean of the Laspeyre's and Paasche's index.

$$P_{on} = \sqrt{\text{Laspeyre's Index} \times \text{Paasche's Index}}$$

$$= \sqrt{\frac{\Sigma p_n q_o}{\Sigma p_o q_o} \times 100 \,.\, \frac{\Sigma p_n q_n}{\Sigma p_o q_n} \times 100} = \sqrt{\frac{\Sigma p_n q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_n q_n}{\Sigma p_o q_n}} \times 100$$

This index lies between Laspeyre's and Paasche's indices.

5.10.2. Weighted Average of Relatives Price Index Numbers

In this method the unweighted index number is converted into weighted index number. The weights used are values. These values are estimated on the basis of the aggregate expenditure in the base year. Weights are in proportion of these values. The aggregate expenditure of a commodity in the base year is calculated by multiplying quantity with price. The index number for the current year is calculated by dividing the sum of the products of the current years price relatives and base years values by the total of the weights. Symbolically

$$\text{Weighted Average of Relatives} = \frac{\Sigma WI}{\Sigma W}$$

where $I = \dfrac{p_n}{p_o} \times 100$ and $W = p_o\, q_o$

Example 5.8.

Construct index numbers of prices from the following data by applying:

(i) Laspeyre's method (ii) Paasche's method and (iii) Fisher's Ideal method

	Base year		Current year	
Commodities	Price	Quantity	Price	Quantity
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

Solution:

The necessary calculations are given below:

	Prices		Quantities					
Commodities	p_o	p_1	q_o	q_1	p_1q_o	p_oq_o	p_1q_1	p_oq_1
A	6	10	50	56	500	300	560	336
B	2	2	100	120	200	200	240	240
C	4	6	60	60	360	240	360	240
D	10	12	30	24	360	300	288	240
E	8	12	40	36	480	320	432	288
					$\Sigma p_1q_o = 1900$	$\Sigma p_oq_o = 1360$	$\Sigma p_1q_1 = 1880$	$\Sigma p_oq_1 = 1344$

(i) Laspeyre's method: $P_{01} = \dfrac{\Sigma p_1q_o}{\Sigma p_oq_o} \times 100 = \dfrac{1900}{1360} \times 100 = 139.71$

(ii) Paasche's method: $P_{01} = \dfrac{\Sigma p_1q_1}{\Sigma p_oq_1} \times 100 = \dfrac{1880}{1344} \times 100 = 139.88$

(iii) Fisher's Ideal method: $P_{01} = \sqrt{\dfrac{\Sigma p_1q_o}{\Sigma p_oq_o} \times \dfrac{\Sigma p_1q_1}{\Sigma p_oq_1}} \times 100$

$$= \sqrt{\frac{1900}{1360} \times \frac{1880}{1344}} \times 100 = 139.79$$

Example 5.9.

Given the following information:

$\Sigma p_1q_o = 4220$, $\Sigma p_oq_o = 3520$, $\Sigma p_2q_0 = 5460$, $\Sigma p_1q_1 = 4810$,

$\Sigma p_0q_1 = 4020$, $\Sigma p_2q_2 = 6896$, $\Sigma p_oq_2 = 4462$.

Compute: (i) Laspeyre's index numbers

(ii) Paasche's index numbers

(iii) Fisher's Ideal index numbers

Solution:

The necessary calculations are given below:

(i) Laspeyre's Price Index Numbers:

$$P_{01} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100 = \frac{4220}{3520} \times 100 = 119.89$$

$$P_{02} = \frac{\Sigma p_2 q_o}{\Sigma p_o q_o} \times 100 = \frac{5460}{3520} \times 100 = 155.11$$

(ii) Paasche's Price Index Numbers:

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_o q_1} \times 100 = \frac{4810}{4020} \times 100 = 119.65$$

$$P_{02} = \frac{\Sigma p_2 q_2}{\Sigma p_o q_2} \times 100 = \frac{6896}{4462} \times 100 = 154.55$$

(iii) Fisher's Ideal Price Index Numbers:

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_1 q_1}{\Sigma p_o q_1}} \times 100 = \sqrt{\frac{4220}{3520} \times \frac{4810}{4020}} \times 100 = 119.77$$

$$P_{02} = \sqrt{\frac{\Sigma p_2 q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_2 q_2}{\Sigma p_o q_2}} \times 100 = \sqrt{\frac{5460}{3520} \times \frac{6896}{4462}} \times 100 = 154.83$$

Example 5.10.

The following data relate to the prices and quantities of three commodities in the years 2002 and 2003. Construct the following index numbers of price for the year 2003 by taking 2002 as the base year.

(i) Base year weighted (ii) Current year weighted.

	Prices (Rs. per kg.)		Quantities (kgs.)	
Commodities	2002	2003	2002	2003
A	6	10	50	60
B	2	2	100	120
C	4	6	60	60

Solution:

The necessary calculations are given below:

	Prices		Quantities					
Commodities	p_o	p_1	q_o	q_1	p_1q_o	p_oq_o	p_1q_1	p_oq_1
A	6	10	50	60	500	300	600	360
B	2	2	100	120	200	200	240	240
C	4	6	60	60	360	240	360	240
					$\Sigma p_1q_o = 1060$	$\Sigma p_oq_o = 740$	$\Sigma p_1q_1 = 1200$	$\Sigma p_oq_1 = 840$

(i) Base year weighted index for 2003,

$$P_{01} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100 = \frac{1060}{740} \times 100 = 143.24$$

(ii) Current year weighted index for 2003,

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_o q_1} \times 100 = \frac{1200}{840} \times 100 = 142.86$$

5.11. CONSUMER PRICE INDEX NUMBERS

A consumer price index number is an index number which is made to measure the relative change in purchasing a specified basket of goods and services between two periods for a certain locality for fixed income group of people. Consumer price index numbers are also called as cost of living index numbers or retail price index numbers. The basket of goods and services will include items like:

(i) food (ii) fuel and lighting (iii) house rent (iv) clothing (v) miscellaneous.

5.12. MAIN STEPS TO CONSTRUCT THE CONSUMER PRICE INDEX NUMBERS

The main steps involved in the construction of consumer price index numbers or cost of living index numbers are given below:

5.12.1. Scope

The first and most important step for constructing a consumer price index number is that we should know the scope of our index number. We should specify the area and locality for which index number is being made and also the income group of the population living in that locality. For example industrial workers, clerks or low salaried workers etc.

5.12.2. Family Budget Enquiry and Allocation of Weights

The next step is to conduct a family budget enquiry. The enquiry should include questions about quality and quantity of goods and services consumed and amount of money spent on them under various items. For example food, fuel and lighting, clothing, house rent and miscellaneous. The miscellaneous group includes items like amusement, education, medicine, gifts, newspaper, transport, barber etc.

5.12.3. Collection of Price Data

The next step is to collect price data for the basket of goods and services for two periods i.e. base year and current year. The price quotations should be taken from those shops which are situated in that locality for which index number is being constructed.

5.12.4. Computation of Consumer Price Index Numbers

There are two methods to compute consumer price index numbers.

(1) AGGREGATE EXPENDITURE METHOD

In this method, the quantities of commodities consumed by the particular group in the base year are estimated and these figures or their proportions are used as weights. Then the total expenditur each commodity for each year is

calculated. The price of the current year is multiplied by the quantity or weight of the base year. These products are added. Similarly for the base year total expenditure on each commodity is calculated by multiplying the quantity consumed by its price in the base year. These products are also added. The total expenditure of the current year is divided by the total expenditure of the base year and the resulting figure is multiplied by 100 to get the required index numbers. In this method, the current period quantities are not used as weights because these quantities change from year to year.

$$P_{on} = \frac{\Sigma p_n q_o}{\Sigma p_o q_o} \times 100$$

where, p_n represent the prices of the current year,

p_o represent the prices of the base year and

q_o represent the quantities consumed in the base year.

(2) FAMILY BUDGET METHOD

In this method, the family budget of a large number of people are carefully studied and the aggregate expenditure of the average family on various items is estimated. These values are used as weights. Current year's prices are converted into price relatives on the basis of base year's prices and these price relatives are multiplied by the respective values of the commodities, in the base year. The total of these products is divided by the sum of the weights and the resulting figure is the required index numbers.

$$P_{on} = \frac{\Sigma WI}{\Sigma W} \quad \text{where} \quad I = \frac{p_n}{p_o} \times 100 \quad \text{and} \quad W = p_o q_o$$

This can be written as

$$P_{on} = \frac{\Sigma \left(\frac{p_n}{p_o}\right) \times 100 (p_o\, q_o)}{\Sigma p_o\, q_o} = \frac{\Sigma p_n\, q_o}{\Sigma p_o\, q_o} \times 100$$

This formula is the same thing as used above in aggregate expenditure method. Thus both the methods i.e., the aggregate expenditure method and family budget method will give the same answer.

Example 5.11.

Compute the cost of living index number by using weighted average of price relatives method.

Items	Prices in rupees		Weights
	Base year	Current year	
A	13	8	6
B	15	22	5
C	249	185	4
D	328	250	1
E	497	448	2

Solution:

The necessary calculations are given below:

Items	Weights (W)	Base year price in Rs. p_o	Current year price in Rs. p_1	$I = \frac{p_1}{p_o} \times 100$	WI
A	6	13	8	$\frac{8}{13} \times 100 = 61.54$	369.24
B	5	15	22	$\frac{22}{15} \times 100 = 146.67$	733.35
C	4	249	185	$\frac{185}{249} \times 100 = 74.30$	297.20
D	1	328	250	$\frac{250}{328} \times 100 = 76.22$	76.22
E	2	497	448	$\frac{448}{497} \times 100 = 90.14$	180.28
	$\Sigma W = 18$				$\Sigma WI = 1656.29$

Cost of living index number (by weighted average of price relatives method)

$$= \frac{\Sigma WI}{\Sigma W} = \frac{1656.29}{18} = 92.02$$

Example 5.12.

An enquiry into the budgets of middle class families in a city gave the following information.

Expenses on	Food 30 %	Rent 15 %	Clothing 20 %	Fuel 10 %	Education 25 %
Prices in 2000 (Rs.)	3000	600	2100	600	1200
Prices in 2001 (Rs.)	3600	600	2400	600	1650

What is the change in the cost of living figure in 2001 as compared with 2000?

Solution:

The necessary calculations are given below:

Commodities	Prices 2000 p_o	Prices 2001 p_1	Weight W	$I = \frac{p_1}{p_o} \times 100$	Product WI
Food	3000	3600	30	$\frac{3600}{3000} \times 100 = 120.00$	3600.0
Rent	600	600	15	$\frac{600}{600} \times 100 = 100.00$	1500.0
Clothing	2100	2400	20	$\frac{2400}{2100} \times 100 = 114.29$	2285.8
Fuel	600	600	10	$\frac{600}{600} \times 100 = 100.00$	1000.0
Education	1200	1650	25	$\frac{1650}{1200} \times 100 = 137.50$	3437.5
			$\Sigma W = 100$	---	$\Sigma WI = 11823.3$

$$\text{Cost of living index number for 2001} = \frac{\Sigma WI}{\Sigma W} = \frac{11823.3}{100} = 118.233 \text{ or } 118.2.$$

Since the cost of living index number for 2001 is greater than 100, we therefore conclude that prices in 2001 have increased 18.2 % as compared with the prices in 2000.

Example 5.13.

Construct the consumer price index number for 2003 on the basis of 2002 from the following data using: (i) Aggregate expenditure method and (ii) Family budget method.

Commodities	Quantity Consumed in 2002	Unit of price	Prices 2002	Prices 2003
A	6 quintal	per quintal	315.75	316.00
B	6 quintal	per quintal	305.00	308.00
C	1 quintal	per quintal	416.00	419.00
D	6 quintal	per quintal	528.00	610.00
E	4 kilogram	per kilogram	12.00	11.50
F	1 quintal	per quintal	1020.00	1015.00

Solution:

(i) Consumer price index number of 2003 by Aggregate expenditure method:

Commodities	Quantity consumed in 2002 q_o	Unit of price	Prices 2002 p_o	Prices 2003 p_1	p_1q_o	p_oq_o
A	6 quintal	per quintal	315.75	316.00	1896	1894.5
B	6 quintal	per quintal	305.00	308.00	1848	1830.0
C	1 quintal	per quintal	416.00	419.00	419	416.0
D	6 quintal	per quintal	528.00	610.00	3660	3168.0
E	4 kilogram	per kilogram	12.00	11.50	46	48.0
F	1 quintal	per quintal	1020.00	1015.00	1015	1020.0
					$\Sigma p_1 q_o = 8884$	$\Sigma p_o q_o = 8376.5$

Consumer price index number of 2003 is:

$$P_{01} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100 = \frac{8884}{8376.5} \times 100 = 106.06$$

(ii) Consumer price index number of 2003 by Family budget method:

Commodities	q_o	Unit of Price	2002 p_o	2003 p_1	$W = p_o q_o$	$I = \frac{p_1}{p_o} \times 100$	Product WI
A	6 quintal	per quintal	315.75	316.00	1894.5	100.08	189601.56
B	6 quintal	per quintal	305.00	308.00	1830.0	100.98	184793.40
C	1 quintal	per quintal	416.00	419.00	416.0	100.72	41899.52
D	6 quintal	per quintal	528.00	610.00	3168.0	115.53	365999.04
E	4 kilogram	per kilogram	12.00	11.50	48.0	95.83	4599.84
F	1 quintal	per quintal	1020.00	1015.00	1020.0	99.51	101500.20
					$\Sigma W = 8376.5$		$\Sigma WI = 888393.56$

Consumer price index number of 2003 is:

$$P_{01} = \frac{\Sigma WI}{\Sigma W} = \frac{888393.56}{8376.5} = 106.06$$

Example 5.14.

Calculate the price index number for 1994 on the basis of 1984 from the following data using aggregate expenditure method.

Commodities	Prices (Rs. per kg.) 1984	Prices (Rs. per kg.) 1994	Values
Wheat	2.50	4.50	100
Rice	9.00	16.00	180
Sugar	8.00	16.00	80
Ghee	12.00	40.00	60

Solution:

Values mean the expenditure on different commodities. Thus values = $p_o q_o$.

The necessary calculations are given below:

Commodities	Prices (Rs.) 1984 p_o	Prices (Rs.) 1994 p_1	$p_o q_o$	$q_o = \frac{p_o q_o}{p_o}$	$p_1 q_o$
Wheat	2.50	4.50	100	$\frac{100}{2.50} = 40$	180
Rice	9.00	16.00	180	$\frac{180}{9.00} = 20$	320
Sugar	8.00	16.00	80	$\frac{80}{8.00} = 10$	160
Ghee	12.00	40.00	60	$\frac{60}{12.00} = 5$	200
			$\Sigma p_o q_o = 420$	--	$\Sigma p_1 q_o = 860$

Price index by Laspeyre's formula:

$$P_{01} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100 = \frac{860}{420} \times 100 = 204.76.$$

Note: From the given data only the Laspeyre's price index can be calculated.

SHORT DEFINITIONS

Index Number

An index number is a measure of change in magnitude relative to a fixed base value.

or

An index number is a ratio used to measure relative change between two time periods.

Simple Index Number

A simple index number is based on the relative changes in the price or quantity of a single commodity. *or*

A simple index measures the relative change from the base period for a single commodity.

Composite Index Number

A composite index number represents combinations of the prices or quantities of several commodities. *or*

A composite index number measures relative change from the base period for a group of items.

Price Index Number

An index that measures a change in prices paid or received by consumers or producers is called a price index number. *or*

An index number constructed to measure the change in the price of an item or a group of items is called a price index number.

Quantity Index Number

An index that is designed to measure changes in quantities with respect to time. *or*

An index number constructed to measure a change in the physical quantity of goods produced, consumed or sold of an item or a group of items is called quantity index number.

Price Relative

The price relative is defined as the price in the current period divided by the price in the base period multiplied by 100 i.e; $P_{on} = \frac{p_n}{p_o} \times 100$.

Fixed Base Method

In fixed base method, one of the time periods is chosen as the base and rest of the values of the various time periods are divided by the base period value and the results are expressed in percentage form.

Link Relative

A link relative is obtained by dividing the current period price by the immediately preceding period price into 100 i.e; $P_{n-1,n} = \frac{p_n}{p_{n-1}} \times 100$.

Chain Base Method

Chain base method is used to convert the link relatives into fixed base method. The process of changing the link relatives into fixed base is known as chaining process. The average of the link relatives for the beginning year is taken as chain index for

that year. This chain index is multiplied by the average of the link relatives of the next year and divided by 100 and so on. The values obtained in this way are known as the chain indices.

Unweighted Index Number

It is a device that measures the relative charge in a group of variables after ignoring the relative importance of the variables.

Simple Aggregative Price Index Number

A simple aggregative price index number is the ratio of a sum of prices in current period to the sum of prices in the base period, expressed as a percentage i.e;

$$P_{on} = \frac{\Sigma p_n}{\Sigma p_o} \times 100.$$ *or*

A simple aggregative price index number is obtained by dividing the sum of current year prices by the sum of base year prices and multiplying by 100.

Weighted Index Number

An index that measures the relative change in a group of variables keeping in view the relative importance of the variables.

Weighted Aggregative Price Index Number

A weighted aggregative price index is the ratio of an aggregate of weighted commodity prices for a given year to an aggregate of the weighted prices of the same commodities in some base year, expressed as a percentage. *or*

A weighted aggregative price index number is the ratio of a weighted sum of prices in current year to a weighted sum of prices in the base year, expressed as a percentage.

Laspeyre's Price Index Number

The Laspeyre's price index number is the ratio of the total cost in the given year of the quantity of each commodity consumed in the base year to the total cost of these quantities in the base year, expressed as a percentage. In symbols, the Laspeyre's index is $P_{on} = \frac{\Sigma p_n q_o}{\Sigma p_o q_o} \times 100.$

Paasche's Price Index Number

The Paasche's price index number is the ratio of the total cost in the given year of the quantity of each commodity consumed in the given year to the total cost of these quantities in the base year, expressed as a percentage. In symbols, the Paasche's index is $P_{on} = \frac{\Sigma p_n q_n}{\Sigma p_o q_n} \times 100$

Consumer Price Index Number or Cost of Living Index Number

A cost of living index number is an index number which is designed to measure the relative change in purchasing a specified basket of goods and services between two periods for a certain locality for fixed income group of people. *or*

A monthly price index that is used to measure the change in prices of basket of goods and services is called consumer price index number.

MULTIPLE CHOICE QUESTIONS

1. An index number is called a simple index when it is computed from:
   (a) single variable (b) bi-variable
   (c) multiple variables (d) none of them
2. Index numbers are expressed in:
   (a) ratios (b) squares
   (c) percentages (d) combinations
3. If all the values are of equal importance, the index numbers are called:
   (a) weighted (b) un-weighted
   (c) composite (d) value index
4. Index numbers can be used for:
   (a) forecasting (b) fixed prices
   (c) different prices (d) constant prices
5. Index for base period is always taken as:
   (a) 100 (b) one
   (c) 200 (d) zero
6. When the prices of rice are to be compared, we compute:
   (a) volume index (b) value index
   (c) price index (d) aggregative index
7. When index number is calculated for several variables, it is called:
   (a) composite index (b) whole sale price index
   (c) volume index (d) simple index
8. How many types are used for the calculation of index numbers:
   (a) 2 (b) 3
   (c) 4 (d) 5
9. In chain base method, the base period is:
   (a) fixed (b) not fixed
   (c) constant (d) zero
10. Which formula is used in chain indices?
    (a) $\frac{\Sigma p_n}{\Sigma p_0} \times 100$ (b) $\frac{p_n}{p_{n-1}} \times 100$
    (c) $\frac{p_n}{p_0} \times 100$ (d) $\frac{\Sigma p_n q_n}{\Sigma p_0 q_0} \times 100$
11. Price relatives is a percentage ratio of current year price and:
    (a) base year quantity (b) previous year quantity
    (c) base year price (d) current year quantity
12. Indices calculated by the chain base method are free from:
    (a) seasonal variations (b) errors
    (c) percentages (d) ratios
13. The chain base indices are not suitable for:
    (a) long range comparisons (b) short range comparisons
    (c) middle range comparisons (d) all of the above

14. An index number that can serve many purposes is called:
    (a) general purpose index (b) special purpose index
    (c) cost of living index (d) none of them
15. Another name of consumer's price index number is:
    (a) whole sale price index number (b) cost of living index
    (c) sensitive (d) composite
16. Consumer price index indicates:
    (a) rise (b) fall
    (c) both (a) and (b) (d) neither (a) and (b)
17. Purchasing power of money can be assessed through:
    (a) simple index (b) Fisher's Ideal index
    (c) consumer price index (d) volume index
18. Cost of living at two different cities can be compared with the help of:
    (a) value index (b) consumer price index
    (c) volume index (d) un-weighted index
19. Consumer price index numbers are obtained by:
    (a) Laspeyre's formula (b) Fisher ideal formula
    (c) Marshall Edgeworth formula (d) Paasche's formula
20. Laspeyre's index = 110, Paasche's index = 108, then Fisher's Ideal index is equal to:
    (a) 110 (b) 108
    (c) 100 (d) 109
21. Most commonly used index number is:
    (a) volume index number (b) value index number
    (c) price index number (d) simple index number
22. For consumer price index, price quotations are collected from:
    (a) fair price shops (b) government depots
    (c) retailers (d) whole-sale dealers
23. Price relatives computed by chain base method are called:
    (a) price relatives (b) chain indices
    (c) link relatives (d) none of them
24. Consumer price index are obtained by:
    (a) Paasche's formula (b) Fisher's ideal formula
    (c) Marshall Edgeworth formula (d) Family budget method formula
25. The aggregative expenditure method and family budget method always give:
    (a) different results (b) approximate results
    (c) same results (d) none of them
26. In fixed base method, the base period should be:
    (a) for away (b) abnormal
    (c) unreliable (d) normal
27. If all the values are not of equal importance the index number is called:
    (a) simple (b) unweighted
    (c) weighted (d) none

28. Which of the following formula satisfy the time reversal test:

(a) $P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$ (b) $P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$

(c) $P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$ (d) none of them

29. When the price of a year is divided by the price of a particular year we get:
    (a) simple relative (b) link relative
    (c) a and b, both (d) none of them
30. When the price of a year is divided by the price of the preceding year, we get:
    (a) value index (b) link relative (c) simple relative (d) none of them
31. The most appropriate average in averaging the price relatives is:
    (a) median (b) harmonic mean
    (c) arithmetic mean (d) geometric mean
32. In constructing index number geometric mean relatives are:
    (a) non-reversible (b) reciprocal
    (c) reversible (d) none of them
33. The general purchasing power of the currency of a country is determined by:
    (a) retail price index (b) volume index
    (c) composite index (d) whole-sale price index
34. What type of index number can help the government to formulate its price policies and to take appropriate economic measures to control prices:
    (a) whole sale price index (b) consumer's price
    (c) quantity (d) none of them
35. The most suitable average in chain base method is:
    (a) arithmetic mean (b) median
    (c) mode (d) geometric mean
36. Base year quantities as weights are used in:
    (a) Laspeyre's method (b) Paasche's method
    (c) Fisher's ideal method (d) difficult to tell
37. Chain process is used to make comparisons of price index numbers in:
    (a) price relative (b) link relative
    (c) simple relative (d) none of the above
38. In the computation of consumer price index numbers, we use:
    (a) aggregate expenditure method (b) family budget method
    (c) chain base method (d) both (a) and (b)
39. The Federal Bureau of Statistics prepares:
    (a) the wholesale price index (b) the consumer price index
    (c) the sensitive price indicator (d) all of the above
40. While computing a weighted index, the current period quantities are used in the:
    (a) Laspeyre's method (b) Paasche's method
    (c) Marshall Edgeworth method (d) Fisher's ideal method

41. The best method to measure the relative change in prices of commodities is:
(a) quantity index number (b) value index number
(c) volume index number (d) price index number

42. When the base year values are used as weights, the weighted average of relatives price index number is the same as the:
(a) Laspeyre's index (b) Paasche's index
(c) simple aggregative index (d) quantity index

43. To measure the relative change in purchasing a specified basket of goods and services between two periods for a certain locality for fixed income group of people, we can use:
(a) consumer price index (b) Paasche's price index
(c) cost of living index (d) both (a) and (c)

44. Fisher's ideal index number is the geometric mean of the:
(a) Laspeyre's and Marshall Edgeworth indices
(b) Laspeyre's and Paasche's indices
(c) Paasche's and Marshall Edgeworth indices
(d) all of the above

45. A number that measures a relative change in a single variable with respect to a base is called:
(a) good index number (b) composite index number
(c) simple index number (d) quantity index number

46. A number that measures an average relative change in a group of related variables with respect to a base is called:
(a) simple index number (b) composite index number
(c) price index number (d) quantity index number

47. An index number constructed to measure the relative change in the price of an item or a group of items is called:
(a) quantity index number (b) price index number
(c) volume index number (d) difficult to tell

48. When relative change is measured for a fixed period, it is called:
(a) chain base method (b) fixed base method
(c) simple aggregative method (d) cost of living method

49. The ratio of a sum of prices in current period to the sum of prices in the base period, expressed as a percentage is called:
(a) simple price index number
(b) simple aggregative price index number
(c) weighted aggregative price index number
(d) quantity index number

50. An index that measures the average relative change in a group of variables keeping in view the relative importance of the variables is called:
(a) simple index number (b) composite index number
(c) weighted index number (d) price index number

51. Link relative of current year is equal to:

(a) $\frac{\text{price of current year}}{\text{price of base year}} \times 100$ (b) $\frac{\text{price of base year}}{\text{price in the preceding year}} \times 100$

(c) $\frac{\text{price in the current year}}{\text{price in the preceding year}} \times 100$ (d) $\frac{\text{price in the preceding year}}{\text{price in the current year}} \times 100$

52. Simple average of relatives is equal to:

(a) $\frac{p_n}{p_o} \times 100$ (b) $\frac{\sum p_n}{\sum p_o} \times 100$

(c) $\sum\left(\frac{p_n}{p_o}\right) \times 100$ (d) $\frac{1}{N}\sum\left(\frac{p_n}{p_o}\right) \times 100$

53. Paasche's price index number is also called:

(a) base year weighted (b) current year weighted

(c) simple aggregative index (d) consumer price index

54. Laspeyre's price index number is also called:

(a) base year weighted (b) current year weighted

(c) cost of living index (d) simple aggregative index

55. Index number having downward bias is:

(a) Laspeyre's index (b) Paasche's index

(c) Fisher's ideal index (d) Marshall Edgeworth index

56. Index number having upward bias is:

(a) Laspeyre's index (b) Paasche's index

(c) Fisher's ideal index (d) Marshall Edgeworth index

57. Marshall Edgeworth price index was proposed by:

(a) one English economist (b) two English economist

(c) three English economist (d) many English economist

58. Index number calculated by Fisher's formula is ideal because it satisfy:

(a) circular test (b) factor reversal test

(c) time reversal test (d) all of the above

59. The test which is not obeyed by any of the weighted index numbers unless the weights are constant:

(a) circular test (b) time reversal test

(c) factor reversal test (d) none of them

Answers

1.	(a)	2.	(c)	3.	(b)	4.	(a)	5.	(a)	6.	(c)	7.	(a)	8.	(a)
9.	(b)	10.	(b)	11.	(c)	12.	(a)	13.	(a)	14.	(a)	15.	(b)	16.	(c)
17.	(c)	18.	(b)	19.	(a)	20.	(d)	21.	(c)	22.	(c)	23.	(c)	24.	(d)
25.	(c)	26.	(d)	27.	(c)	28.	(c)	29.	(a)	30.	(b)	31.	(d)	32.	(c)
33.	(d)	34.	(b)	35.	(d)	36.	(a)	37.	(b)	38.	(d)	39.	(d)	40.	(b)
41.	(d)	42.	(a)	43.	(d)	44.	(b)	45.	(c)	46.	(b)	47.	(b)	48.	(b)
49.	(b)	50.	(c)	51.	(c)	52.	(d)	53.	(b)	54.	(a)	55.	(a)	56.	(b)
57.	(b)	58.	(d)	59.	(a)										

SHORT QUESTIONS

Q.1 Define an index number.
Q.2 Define price index numbers.
Q.3 Define simple index number.
Q.4 Define composite index number.
Q.5 Define weighted index number.
Q.6 Define un-weighted index number.
Q.7 Define price relative.
Q.8 Define link relative.
Q.9 Why index numbers are called economics barometer?
Q.10 Differentiate between simple and composite index numbers.
Q.11 Differentiate between price relatives and link relatives.
Q.12 Describe the different types of index numbers.
Q.13 What are the important uses of index numbers?
Q.14 Explain the fixed base method.
Q.15 Describe the chain base method.
Q.16 Distinguish between fixed base and chain base methods.
Q.17 Write down the main steps involved in the construction of price index numbers.
Q.18 Distinguish between weighted and un-weighted index numbers.
Q.19 Write down the main limitations of index numbers.
Q.20 What are the various ways of assigning weights in the construction of index numbers?
Q.21 Define index numbers. Write down only main steps which are used for the construction of index numbers of prices.
Q.22 Define simple aggregative price index number.
Q.23 Name the sources of index numbers.
Q.24 Define the base period or reference period.
Q.25 Differentiate between implicit weights and explicit weights.
Q.26 Define Laspeyre's price index number or base year weighted.
Q.27 Define Paasche's price index number or current year weighted.
Q.28 Write down the methods of calculating weighted index numbers.
Q.29 Define weighted aggregative price index number.
Q.30 What are the two methods of constructing un-weighted index numbers?
Q.31 Explain the meaning of consumer price index numbers.
Q.32 Explain the difference between wholesale price and consumer price index numbers.
Q.33 What is cost of living index number?
Q.34 Write down the methods to compute consumer price index numbers.
Q.35 Write down the main steps to construct the consumer price index numbers.
Q.36. What is CPI? How is it calculated?
Q.37 Write short note on quantity index number.
Q.38 What are the different types of index numbers?

Q.39 Explain a market basket.

Q.40 What is the relationship between Laspeyre's price index number, Paasche's price index number and Fisher's Ideal price index number?

Q.41 How we find weighted aggregative price index numbers?

Q.42 How we compute link relatives in chain base method?

Q.43 How we find a value index number?

Q.44 How we find simple aggregative price index number?

Q.45 How we find Fisher's Ideal price index number?

Q.46 How we calculate price relatives in fixed base method?

Q.47 How we calculate un-weighted index numbers?

Q.48 How many methods are used for selection of base period? Name these methods.

Q.49 How we calculate consumer price index numbers?

Q.50 How we compute weighted index numbers?

Q.51 How we compute base year weighted index number of prices?

Q.52 How we calculate current year weighted index for prices?

Q.53 Write down the advantages of chain base method.

Q.54 Given p_0 = 16, 15, 12, 20, 22 and p_1 = 20, 12, 15, 22, 22.

Find I where $I = \frac{p_1}{p_0} \times 100$.

Ans. 125, 80, 125, 110, 100

Q.55 Given p_0 = 6, 2, 4 and q_0 = 50, 100, 60. Find $\sum W$.

Ans. 740

Q.56 Given p_0 = 2, 4, 6, q_0 = 10, 12, 14 and I = 150, 125, 140. Compute $\sum WI$

Ans. 20760

Q.57 Given W = 100, 180, 80, 60 and p_0 = 2, 9, 8, 12. Find q_0.

Ans. 50, 20, 10, 5.

Q.58 Given $\Sigma p_0 = 2550$ and $\Sigma p_n = 2590$. Find price index number by using simple aggregative method.

Ans. 101.57

Q.59 Given $\sum p_0 = 660$, $\sum p_1 = 924$ and $\sum p_2 = 1056$. Compute simple aggregative price index number.

Ans. $P_{01} = 140$, $P_{02} = 160$

Q.60 If link relatives are 100, 120, 102, 105, 118 and 112. Find chain indices.

Ans. 100, 120, 122.4, 128.52, 151.65, 169.85

Q.61 Given $\Sigma p_0 q_0 = 850$ and $\Sigma p_1 q_0 = 1170$. Find Laspeyre's price index number.

Ans. 137.65

Q.62 Given $\sum p_1 q_0 = 1050$, $\sum p_2 q_0 = 1120$ and $\sum p_0 q_0 = 850$. Compute base year weighted index.

Ans. $P_{01} = 123.53$, $P_{02} = 131.76$

Q.63 Given $\Sigma p_0q_1 = 950$ and $\Sigma p_1q_1 = 1310$. Find Paasche's price index number.

Ans. 137.89

Q.64 Given $\sum p_1q_1 = 1400$, $\sum p_2q_o = 1600$, $\sum p_0q_1 = 1360$ and $\sum p_o q_2 = 1560$. Compute current year weighted index numbers.

Ans. $P_{01} = 102.94$, $P_{02} = 102.56$

Q.65 Given Laspeyre's price index number = 120 and Paasche's price index number = 119.6. Find Fisher's Ideal price index number.

Ans. 119.8

Q.66 Given $\Sigma p_0q_0 = 3600$, $\Sigma p_1q_0 = 4300$, $\Sigma p_0q_1 = 4100$ and $\Sigma p_1q_1 = 4890$. Find Fisher's Ideal price index number.

Ans. 119.36

Q.67 Given $\Sigma p_0q_0 = 1500$ and $\Sigma p_nq_0 = 2040$. Find base year weighted index.

Ans. 136

Q.68 Given $\Sigma p_0q_n = 1000$ and $\Sigma p_nq_n = 1360$. Find current year weighted index.

Ans. 136

Q.69 Given Fisher's Ideal price index number = 104.3 and Paasche's price index number = 103.2. Find Laspeyre's price index number.

Ans. 105.4

Q.70 Given Laspeyre's price index number = 110 and Fisher's Ideal price index number = 111. Find Paasche's price index number.

Ans. 112

Q.71 Given $\Sigma W = 20$ and $\Sigma WI = 1800$. Find cost of living index number by weighted average of price relatives method.

Ans. 90

Q.72 Given $\Sigma W = \Sigma p_0q_0 = 8500$ and $\Sigma WI = 892500$. Construct consumer price index number by family budget method.

Ans. 105

Q.73 Given W = 45, 15, 12, 8, 20 and $\sum WI = 10500$. Compute cost of living index number.

Ans. 105

Q.74 Given W = 20, 25, 30, 40 and I = 100, 105, 110, 120. Find consumer price index number.

Ans. 110.65

Q.75 Given $\Sigma p_1q_0 = 9000$ and $\Sigma p_0q_0 = 8490$. Find and interpret consumer price index number by aggregative expenditure method.

Ans. 106

Q.76 Given N(number of commodities) = 5 and $\Sigma\left(\frac{p_n}{p_0}\right) \times 100 = 510$. Find price index by average of price relatives method.

Ans. 102

EXERCISES

Q.1 Find price relatives for the data given below, using (i) 1992 as base (ii) average of last 3 years as base:

Years	1992	1993	1994	1995	1996	1997	1998	1999	2000
Prices	12.5	15.0	17.0	20.0	25.0	22.5	27.5	37.5	35.0

Ans. (i) 100, 120, 136, 160, 200, 180, 220, 300, 280

(ii) 37.50, 45.00, 51.01, 60.01, 75.01, 67.51, 82.51, 112.51, 105.01

Q.2 Find index numbers of prices from the following data using the average of last five years as base.

Years	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Prices	10.75	12.10	13.10	13.85	14.25	15.90	16.80	14.15	13.20	14.00	10.75

Ans. 78.01, 87.81, 95.07, 100.51, 103.41, 115.38, 121.92, 102.69, 95.79, 101.60, 78.01.

Q.3 Compute the index numbers for each year from the following average wholesale prices of cotton in rupees per bale of 10 lbs. for the period 1980 to 1990 taking 1980 = 100.

Years	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Prices (Rs.)	65	72	75	76	80	85	82	88	88	90	95

Ans. 100, 110.77, 115.38, 116.92, 123.08, 130.77, 126.15, 135.38, 135.38, 138.46, 146.15

Q.4 Table given below shows a country's average wholesale prices of wheat for various years. Find the price relatives for (i) the year 1958 using 1948 as base (ii) the years 1949 and 1956 using 1950 as base (iii) the years 1955 – 1958 using 1947 – 1949 = 100.

Years	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958
Prices	2.66	2.50	2.24	2.29	2.41	2.45	2.49	2.56	2.50	2.39	2.35	2.23

Ans. (i) 89.2 (ii) 97.82 104.37 (iii) 101.21, 96.76, 95.14, 90.28

Q.5 The average retail prices in dollars per ton of wheat in a certain country during the years 1953 – 1958 are given. (i) using 1953 as base, find the price relatives corresponding to the years 1956 and 1958. (ii) using 1956 as base, find the price relatives corresponding to all the given years. (iii) using 1953 – 1955 as base, find the price relatives corresponding to all the given years.

Years	1953	1954	1955	1956	1957	1958
Average retail price	14.95	14.94	15.10	15.65	16.28	16.53

Ans. (i) 104.68, 110.57 (ii) 95.53, 95.46, 96.49, 100, 104.03, 105.62

(iii) 99.67, 99.60, 100.67, 104.33, 108.53, 110.20

Q.6 Compute chain index numbers for the following data taking 1997 as base year.

Year	1997	1998	1999	2000	2001	2002	2003
Prices	180	185	194	200	204	218	220

Ans: 100, 102.78, 107.78, 111.11, 113.33, 121.10, 122.21

Q.7 Convert the following prices into price relatives using chain base method taking 1994 as base year.

Year	1994	1995	1996	1997	1998	1999	2000
Prices	36	45	56	64	70	80	90

Ans: 100, 125, 155.55, 177.78, 194.44, 222.23, 250.01

Q.8 From the fixed base index numbers given below, prepare chain base index numbers taking 1990 as base year.

Years	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Fixed indices	120	116	120	120	137	136	149	156	137	162	149

Ans. 100, 96.67, 103.45, 100, 114.17, 99.27, 109.56, 104.70, 87.82, 118.25, 91.98

Q.9 The following are the prices of six different commodities for 1989 and 1990. Compute a price index by (i) simple aggregative method and (ii) average of price relative method by using both arithmetic mean and geometric mean, taking 1989 as base.

Commodity	A	B	C	D	E	F
Price in 1989 (Rs.)	40	60	20	50	80	100
Price in 1990 (Rs.)	50	60	30	70	90	110

Ans. (i) 117.14 (ii) 122.92, 121.70.

Q.10 Compute index numbers of prices from the following data taking 1981 as base and using median as an average:

	Prices		
Year	A	B	C
1981	18	85	52
1982	22	76	60
1983	28	80	66
1984	31	95	80

Ans. 100, 115.38, 126.92, 153.85,

Q.11 Given the prices of four commodities, construct price index numbers by using simple aggregative method taking (i) 1962 as base (ii) average of all years aggregate as base.

	Commodities (prices in Rs.)			
Year	Firewood	Soft Cake	Oil	Matches
1962	3.25	2.50	0.20	0.06
1963	3.44	2.80	0.22	0.06
1964	3.50	2.00	0.25	0.06
1965	3.75	2.50	0.25	0.06

Ans. (i) 100, 108.49, 96.67, 109.15 (ii) 96.55, 104.74, 93.33, 105.38.

Q.12 Construct chain base indices for the following years taking 1940 as base:

Year	1940	1941	1942	1943	1944
Wheat	2.8	3.4	3.6	4.0	4.2
Rice	10.5	10.8	10.6	11.0	11.5
Maize	2.7	3.2	3.5	3.8	4.0

Ans. 100, 114.27, 119.38, 128.72, 135.08

Q.13 Use chain base method to construct index numbers:

Items	1958	1960	1961
A	10	12	13.5
B	16	16.5	17
C	18	18.5	19
D	20	21	21.5

Ans. 100, 107.73, 113.28

Q.14 Compute the chain indices from the following price relatives:

Years	Sugar	Tea	Coffee
1921	81	82	111
1922	77	96	119
1923	87	88	128
1924	75	89	139
1925	90	84	146

Ans. 91.33, 97.22, 101.18, 99.81, 106.27

Q.15 Given the price relatives of three commodities, construct the chain indices using median as an average:

	Price relatives		
Year	Wheat	Rice	Sugar
1984	100	100	100
1985	105	97	125
1986	110	100	130
1987	110	99	128
1988	120	105	130

Ans. 100, 105, 109.2, 108.11, 114.66

Q.16 Find the chain indices from the following price relatives of three commodities using the geometric mean.

	Commodities		
Year	A	B	C
1996	81	77	119
1997	62	54	128
1998	104	87	111
1999	93	75	154
2000	60	43	165

Ans. 90.54, 75.39, 100.14, 102.41, 75.22

Q.17 Compute the index numbers using simple aggregative method with 1952 as base year.

Commodity	1952	1953	1954	1955
Wheat	25.2	21.3	25.4	30.2
Rice	15.9	16.3	18.9	19.3
Barley	15.9	14.0	16.3	18.5
Jawar	11.3	14.3	11.5	13.6
Grams	13.0	13.5	13.6	13.9

Ans: 100, 97.66, 105.41, 117.47

Q.18 Given the prices of three commodities, compute price index numbers by simple aggregative method taking 1998 as base.

Year	Commodities A	B	C
1998	75	62	100
1999	52	44	118
2000	94	77	101
2001	83	65	144
2002	50	48	155

Ans: 100, 90.30, 114.77, 123.21, 106.75

Q.19 Compute chain index numbers from the following price relatives using geometric mean as an average and taking 2000 as base year.

Year	Commodities A	B	C
2000	4	10	20
2001	4	6	23
2002	7	9	25
2003	8	10	30

Ans: 100, 88.37, 125.34, 144.23

Q.20 Construct the following weighted aggregative index numbers of prices for the year 1981 from the data given below:

(i) Laspeyre's index number (ii) Paasohe's index number

(iii) Fisher's index number.

Commodity	Prices 1980 (base)	Prices 1981	Quantities 1980 (base)	Quantities 1981
A	10	12	20	22
B	8	8	16	18
C	5	6	10	11
D	4	5	7	8

Ans. (i) 114.04 (ii) 113.97 (iii) 114

Q.21 Given the following information:

	2002		2003	
Commodity	Price	Quantity	Price	Quantity
C_1	45	90	93	100
C_2	37	10	64	11
C_3	27	3	51	5

Construction the following index numbers of prices for the year 2003 by taking 2002 as the base year.

(i) Base year weighted (ii) Current year weighted.

Ans. (i) 203.58 (ii) 203.47

Q.22 Construct price index numbers for the year 1992 on the basis of the year 1987 of the following by using Laspeyre's, Paasche's and Fisher's Ideal index number formulae:

	A		B		C	
Year	Price	Quantity	Price	Quantity	Price	Quantity
1987	5	10	8	26	6	13
1992	4	12	7	27	5	14

Ans. 85.42, 85.28, 85.35

Q.23 Compute the index numbers by Fisher's Ideal formula for the data given below:

$\Sigma p_1q_0 = 8800$, $\Sigma p_0q_0 = 7800$, $\Sigma p_1q_1 = 15400$, $\Sigma p_0q_1 = 13710$,
$\Sigma p_2q_0 = 11000$, $\Sigma p_2q_2 = 23000$, $\Sigma p_0q_2 = 16370$

Ans. 112.57, 140.76

Q.24 From the following data, find index numbers for 2002 with 2001 as base-year by (i) Laspeyre's formula (ii) Paasche's formula. (iii) show numerically that Fisher's Ideal formula is the geometric mean of these two:

	Prices		Quantities	
Items	2001	2002	2001	2002
A	64	75	270	290
B	40	45	124	144
C	18	21	130	137
D	58	68	185	200

Ans. (i) 116.51 (ii) 116.47 (iii) 116.49

Q.25 The following data gives the prices and quantities of four commodities for the years 2000 and 2002.

	Prices		Quantities	
Commodity	2000	2002	2000	2002
A	70	75	300	310
B	72	80	240	275
C	25	32	132	148
D	60	85	280	360

Construct the index number of prices for the year 2002 taking 2000 as base year using Paasche's formula.

Ans. 120.64

Q.26 Given the following information:

$\Sigma p_1q_0 = 41140$, $\Sigma p_0q_0 = 35310$, $\Sigma p_1q_1 = 46707$, $\Sigma p_0q_1 = 40048$,

$\Sigma p_2q_0 = 39644$, $\Sigma p_2q_2 = 51724$, $\Sigma p_0q_2 = 47376$

Compute: (i) Base year weighted price index

(ii) Current year weighted price index

(iii) Fisher's Ideal price index

Ans. (i) 116.51, 112.27 (ii) 116.63, 109.18 (iii) 116.57, 110.72

Q.27 An enquiry into the budgets of the middle class families of a certain city revealed that on average the percentage expenses on the different groups were food 45, rent 15, clothing 12, fuel and light 8 and miscellaneous 20. The group index numbers for the current year as compared with a fixed base period were respectively 410, 150, 343, 248 and 285. Calculate the consumer price index number for the current year.

Ans. 325

Q.28 Compute the cost of living index number by using weighted average of price relatives method.

Commodity	Weights	Base year price in Rs.	Current year price in Rs.
A	4	2	7
B	1	5	2
C	5	8	5
D	2	5	12
E	3	3	6

Ans. 188.83

Q.29 From the data given below construct consumers price index numbers of 1986 on the basis of 1976 by using (i) aggregative expenditure method (ii) family budget method. Test whether the answer is the same by both the methods.

Food	Prices		Quantity consumed
	1976	1986	1976
Wheat	8	14	4
Rice	15	21	2
Dal	10	14	1
Oil	20	30	5
Ghee	6	12	3
Cereals	7	14	1
Vegetables	5	15	2

Ans. (i) 165.22 (ii) 165.22

Q.30 Calculate the index number for 1990 on the basis of 1978.

Commodity	Unit of price	Quantity consumed	Prices	
			1978	1990
Wheat	Rs. per maund	20 seers	10.00	13.50
Dal	Rs. per maund	8 seers	15.00	20.00
Oil	Rs. per maund	$1\frac{1}{2}$ seers	90.25	200.50
Fuel	Rs. per maund	4 maunds	2.25	2.50
Clothing	Rs. per yard	22 yards	1.50	2.25

Ans. 145.67

Q.31 Construct consumers price index numbers of 1980 on the basis of 1978 using:

(i) aggregative method (ii) family budget method:

Articles	Quantity consumed in 1978	Unit of price	Prices 1978	Prices 1980
Rice	6 maunds	per seer	6.00	6.50
Wheat	10 maunds	per maund	35.00	40.00
Grain	3 maunds	per maund	60.00	90.00
Pulses	5 maunds	per maund	120.00	144.00
Ghee	5 seers	per seer	8.00	10.00
Sugar	1 maund	per maund	240.00	300.00

Ans. (i) 115.79 (ii) 115.79.

Q.32 An inquiry into the budgets of middle class families in a certain city gave the following information:

Expenses on	Food 35 %	Rent 15 %	Clothing 20 %	Fuel 10 %	Miscellaneous 20 %
Prices (1989) Rs.	1500	500	1000	200	600
Prices (1990) Rs.	1740	600	1250	250	900

What changes are observed in the cost of living index number of 1990 as compared with that of 1989?

Ans. 126.1.

Chapter 6

SET THEORY

6.1. INTRODUCTION

Set theory which was developed by George Cantor (1845 – 1918) between 1874 – 1895, is a basic mathematical tool that is used by various branches of Mathematics, such as probability theory, calculus and geometry. Cantor was born in Russia in 1845, but moved to Germany in 1856. In 1863, he entered the university of Berlin, where he studied under K. Weierstrass (1815 – 1897), L. Kronecker (1823 – 1891), and E.E. Kummer (1810 – 1893), all of whom are famous mathematicians, and received his Ph. D. in 1867. One of the subsequent main works in set theory is by F. Hausdorff, which was first published in 1914.

6.2. SET

A set is any well defined collection of distinct objects. The objects of the set are called its members or elements. We shall denote the set 'A' whose elements are a_1, $a_2, a_3, \cdots, a_n$ by $A = \{ a_1, a_2, a_3, \cdots, a_n \}$. Sets are usually denoted by the capital letters A, B, C or X, Y, Z etc. and the lower case letters a, b, c and x, y, z etc. are used to denote the elements. Sets are always written within pair of brackets. For example the sets A and B are written as $A = \{ 1, 2, 3 \}$, $B = \{$ Aslam, Anwar, Ali $\}$. The element 1 is member of A is written as $1 \in A$ (1 belongs to the set A), Similarly Aslam $\in$ B, Anwar $\in$ B.

A set can be defined by actually listing all its elements or by describing some property of all the members. The first is called tabular form or roster method and the second is called property method. For example, a set V of all vowels in English alphabet can be written by tabular form as $V = \{ a, e, i, o, u \}$ and by property method it can be written as $V = \{ x \mid x \text{ is a vowel} \}$. It is read as the set of all elements x such that x is a vowel. The line / is read as "such that".

6.2.1. NULL SET (EMPTY SET OR VOID SET)

A set that contains no element is defined as the null set and is usually denoted by ϕ. i.e., $\phi = \{ \quad \}$. It is also called empty set or void set.

6.2.2. SINGLETON SET

A set which has only one element is called singleton set. For example, $A = \{$ Mangla Dam $\}$, $B = \{$ Punjab $\}$, $C = \{$ The Ravi $\}$.

6.2.3. FINITE AND INFINITE SETS

A set is finite if it contains a specific number of different elements, i.e. while counting the members of the set the counting process comes to an end. Otherwise a set is infinite. Examples of finite sets are:

$A = \{1, 2, 5\}$ $B = \{x, y, z, t, w, v\}$

$C = \{x / x \text{ is a river on the earth}\}$ $D = \{y / y \text{ is a month of the year}\}$

$E = \{z / z \text{ is a vowel in English alphabet}\}$

Examples of infinite sets are:

$A = \{2, 4, 6, 8, \ldots\}$ $B = \{x / 0 \leq x \leq 2\}$

$C = \{y / y \text{ is a star in the sky}\}$ $D = \{z / z \text{ is a point on a line}\}$

$E = \{w / w \text{ is an odd number}\}$

6.2.4. SUBSET

If every element of a set A is also an element of a set B, then A is called subset of B, and we shall write $A \subseteq B$ or $B \supseteq A$; read as "A is contained in B" or "B contains A".

If A is subset of B and B is subset of A, then we write

$A \subset B \quad B \subset A \quad \text{or} \quad A = B$

If A is subset of B, but A is not equal to B, we say that A is a proper subset of B. For example, if $A = \{a, i, u\}$ and $B = \{a, e, i, o, u\}$, then A is proper subset of B.

6.2.5. UNIVERSAL SET OR SPACE SET

A set which contains all elements of all sets under consideration is called universal set or space set and is denoted by U or S.

If $A = \{1, 2, 3\}$, $B = \{3, 4, 5, 6\}$, $C = \{10, 8\}$, $D = \{9, 6\}$

then $U = S = \{1, 2, 3, 4, 5, 6, 8, 9, 10\}$.

6.2.6. IDENTICAL SETS OR EQUAL SETS

Two sets A and B are said to be identical or equal if they contain exactly the same elements.

If $A = \{1, 2, 3, 4\}$, $B = \{4, 1, 2, 3\}$, A and B are called equal sets.

6.2.7. VENN DIAGRAM

Venn diagram is a diagram in which universal set S is represented by a rectangle and sets are represented by circles or parts of circular regions.

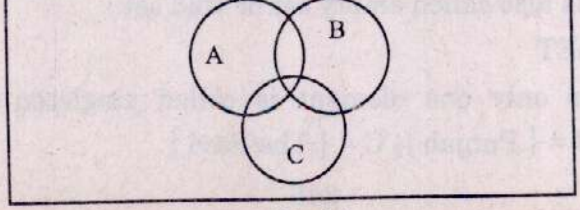


Fig. 6.1.

6.3. OPERATIONS ON SETS

Given a universal set S and its subsets, A, B, C, ..., we can perform on these subsets certain operations that produce other, or perhaps the same, subsets. Among all the operations on sets, we are interested only in union, intersection, difference and complementation.

6.3.1. UNION OF SETS

Union of the two sets A and B is denoted by $A \cup B$. If A and B are two sets, then $A \cup B$ is a set which contains those elements which belong to A or B, or both, i.e.,

$$A \cup B = \{x / x \in A \ \text{ or } \ x \in B\}$$

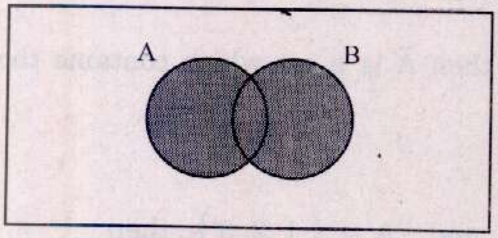
Let $A = \{1, 2, 3\}$
and $B = \{3, 4, 5\}$
then the union of A and B is

$$A \cup B = \{1, 2, 3, 4, 5\}$$

Let $A = \{2, 4, 6\}$
and $B = \{8, 9, 10\}$
then the union of A and B is

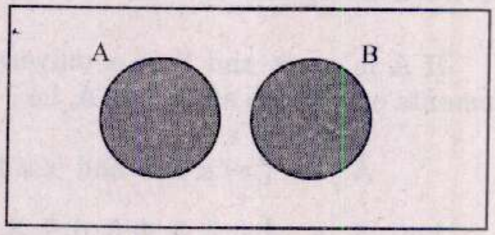
$$A \cup B = \{2, 4, 6, 8, 9, 10\}$$

Venn diagram



$A \cup B$ is shaded area

Fig.6.2.(a)



$A \cup B$ is shaded area

Fig.6.2.(b)

6.3.2. INTERSECTION OF SETS

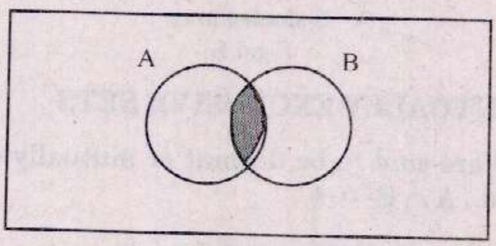
Intersection of the sets A and B is denoted by $A \cap B$. If A and B are two sets, then $A \cap B$ is a set which contains those elements which are common to both A and B, i.e.,

$$A \cap B = \{x / x \in A \ \text{ and } \ x \in B\}$$

Let $A = \{1, 3, 4, 5\}$ and $B = \{2, 3, 4, 6\}$, then the intersection of A and B is:

$$A \cap B = \{3, 4\}$$

Venn diagram



$A \cap B$ is shaded area

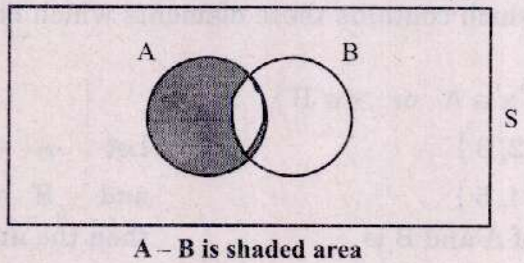
Fig.6.3.

6.3.3. DIFFERENCE OF TWO SETS

If A and B are two sets, then $A - B$ is a set which contains those elements of A which are not in B, i.e., $A - B = \{ x / x \in A \text{ and } x \notin B \}$

Let $A = \{1, 2, 3, 4\}$ and $B = \{3, 4, 5, 6\}$, then $A - B = \{1, 2\}$

Venn diagram



A – B is shaded area

Fig.6.4.

6.3.4. COMPLEMENT SET

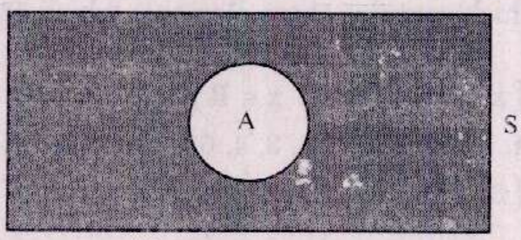
If A is a set and S is a universal set, then $\bar{A}$ is a set which contains those elements of S which are not in A, i.e.,

$$\bar{A} = \{ x / x \in S \text{ and } x \notin A \}$$

Let $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $A = \{1, 2, 3\}$, then

$$\bar{A} = \{4, 5, 6, 7, 8, 9, 10\}$$

Venn diagram



$\bar{A}$ is shaded area

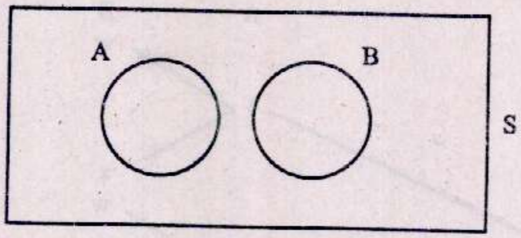
Fig.6.5.

6.4. DISJOINT OR MUTUALLY EXCLUSIVE SETS

Two sets A and B are said to be disjoint or mutually exclusive if they have no elements in common, i.e., $A \cap B = \phi$.

Let $A = \{1, 2, 3, 4\}$ and $B = \{7, 8, 9, 10\}$, then $A \cap B = \phi$.

Venn diagram



$A \cap B = \phi$

Fig.6.6.

6.5. CLASS OF SETS

If the elements of a set E are also sets, then the set E is called a class of sets or a set of sets is called a class of sets.

Let $A = \{1, 2\}$, $B = \{2, 3, 4, 5\}$, $C = \{1, 9\}$ and $D = \{6, 7, 8, 10\}$

and $E = \{A, B, C, D\}$, then E is called class of sets.

6.6. POWER SET

The set of all possible subsets of a set 'A' is called power set of A and is denoted by P(A).

Let $A = \{1, 2, 3\}$

There are $2^3 = 8$ subsets from a set of 3 elements, i.e.,

$A_1 = \{1\}$, $A_2 = \{2\}$,

$A_3 = \{3\}$, $A_4 = \{1, 2\}$, $A_5 = \{1, 3\}$

$A_6 = \{2, 3\}$, $A_7 = \{1, 2, 3\}$ and $A_8 = \{\ \} = \phi$

Then $P(A) = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$

6.7. PRODUCT SET

If A and B are two sets, then $A \times B$ contains all those elements of the type (x, y) where $x \in A$ and $y \in B$, i.e.,

$$A \times B = \{(x, y) / x \in A, y \in B\}$$

Let $A = \{1, 2, 3\}$ and $B = \{w, x\}$, then

$$A \times B = \{(1, w), (1, x), (2, w), (2, x), (3, w), (3, x)\}$$

Tree diagram

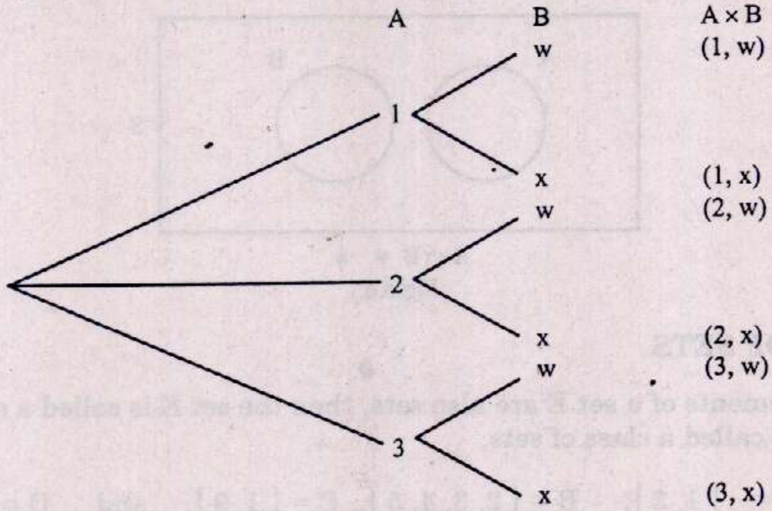


Fig.6.7.

In general $A \times B$ is not equal to $B \times A$.

6.8. PARTITION OF SET

If S is a universal set, and we divide S into different disjoint and non empty subsets, it is called partition of set S.

$$S = \{1, 2, 3, 4\}, \quad A_1 = \{1\}, \quad A_2 = \{2, 3\}, \quad A_3 = \{4\}$$

A_1, A_2 and A_3 form partition of S, i.e.,

(i) $A_i \cap A_j = \phi$ for all $i \neq j$

(ii) $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n = S$

Example 6.1.

If $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

Consider the subsets of S given as:

$$A = \{1, 2, 3, 10\}, B = \{2, 4, 6, 8\}, C = \{1, 3, 5, 7\} \text{ and } D = \{8, 9, 10\}.$$

Then find (i) $A \cup B$ (ii) $A \cup C$ (iii) $A \cap C$ (iv) $C \cap D$

(v) $B \cap D$ (vi) $\bar{D}$ (vii) $\bar{C}$.

Solution:

(i) $A \cup B = \{1, 2, 3, 4, 6, 8, 10\}$ (ii) $A \cup C = \{1, 2, 3, 5, 7, 10\}$

(iii) $A \cap C = \{1, 3\}$ (iv) $C \cap D = \phi$

(v) $B \cap D = \{8\}$ (vi) $\bar{D} = \{1, 2, 3, 4, 5, 6, 7\}$

(vii) $\bar{C} = \{2, 4, 6, 8, 9, 10\}$

Example 6.2.

Let $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{1, 2, 3, 4\}$,
$B = \{2, 4, 6, 8\}$ and $C = \{3, 4, 5, 6\}$.

Find (i) $\bar{A}$ (ii) $\bar{D}$ (iii) $(\overline{A \cap C})$ (iv) $(\overline{A \cup B})$ (v) $(\bar{\bar{A}})$ (vi) $(\overline{B - C})$

Solution:

(i) $\bar{A} = \{5, 6, 7, 8, 9\}$

(ii) $\bar{B} = \{1, 3, 5, 7, 9\}$

$A \cap C = \{3, 4\}$

(iii) $(\overline{A \cap C}) = \{1, 2, 5, 6, 7, 8, 9\}$

$A \cup B = \{1, 2, 3, 4, 6, 8\}$

(iv) $(\overline{A \cup B}) = \{5, 7, 9\}$

(v) $(\bar{\bar{A}}) = \{1, 2, 3, 4\} = A$

$(B - C) = \{2, 8\}$

(vi) $(\overline{B - C}) = \{1, 3, 4, 5, 6, 7, 9\}$

Example 6.3.

Let $A = \{2, 3\}$, $B = \{3, 4\}$ and $C = \{4\}$ be subsets of the universal set $S = \{2, 3, 4\}$. Determine the sets.

(i) $A \times A$ (ii) $A \times B$ (iii) $B \times A$ (iv) $(A \times B) \cap (B \times C)$
(v) $(A \times B) \cup (B \times C)$

Solution:

(i) $A \times A = \{(2, 2), (2, 3), (3, 2), (3, 3)\}$

(ii) $A \times B = \{(2, 3), (2, 4), (3, 3), (3, 4)\}$

(iii) $B \times A = \{(3, 2), (3, 3), (4, 2), (4, 3)\}$

$B \times C = \{(3, 4), (4, 4)\}$

(iv) $(A \times B) \cap (B \times C) = \{3, 4\}$

(v) $(A \times B) \cup (B \times C) = \{(2, 3), (2, 4), (3, 3), (3, 4), (4, 4)\}$

Example 6.4.

Given $S = \{k, l, m, n, o, p\}$, $A = \{k, l\}$, $B = \{m, n, o, p\}$, $C = \{o, p\}$.

Find (i) $\bar{C}$ (ii) $A \cup C$ (iii) $A \cup B$ (iv) $A \cap B$

(v) $A \cup S$ (vi) $S \cap B$ (vii) $(\overline{A \cup B})$ (viii) $(\overline{\bar{A} \cap \bar{B}})$

Solution:

(i) $\bar{C} = \{k, l, m, n\}$ (ii) $A \cup C = \{k, l, o, p\}$

(iii) $A \cup B = \{k, l, m, n, o, p\}$ (iv) $A \cap B = \phi$

(v) $A \cup S = \{k, l, m, n, o, p\} = S$ (vi) $S \cap B = \{m, n, o, p\} = B$

(vii) $(\overline{A \cup B}) = \phi$

(viii) $\bar{A} = \{m, n, o, p\}$ $\bar{B} = \{k, l\}$ $\bar{A} \cap \bar{B} = \phi$

$(\overline{\bar{A} \cap \bar{B}}) = \{k, l, m, n, o, p\} = S$

Example 6.5.

Let $A = \{1, 2, 3\}$, $B = \{2, 4\}$ and $C = \{3, 4, 5\}$. Find $A \times B \times C$.

Solution:

A convenient method of finding $A \times B \times C$ is through the so called "tree diagram" shown below:

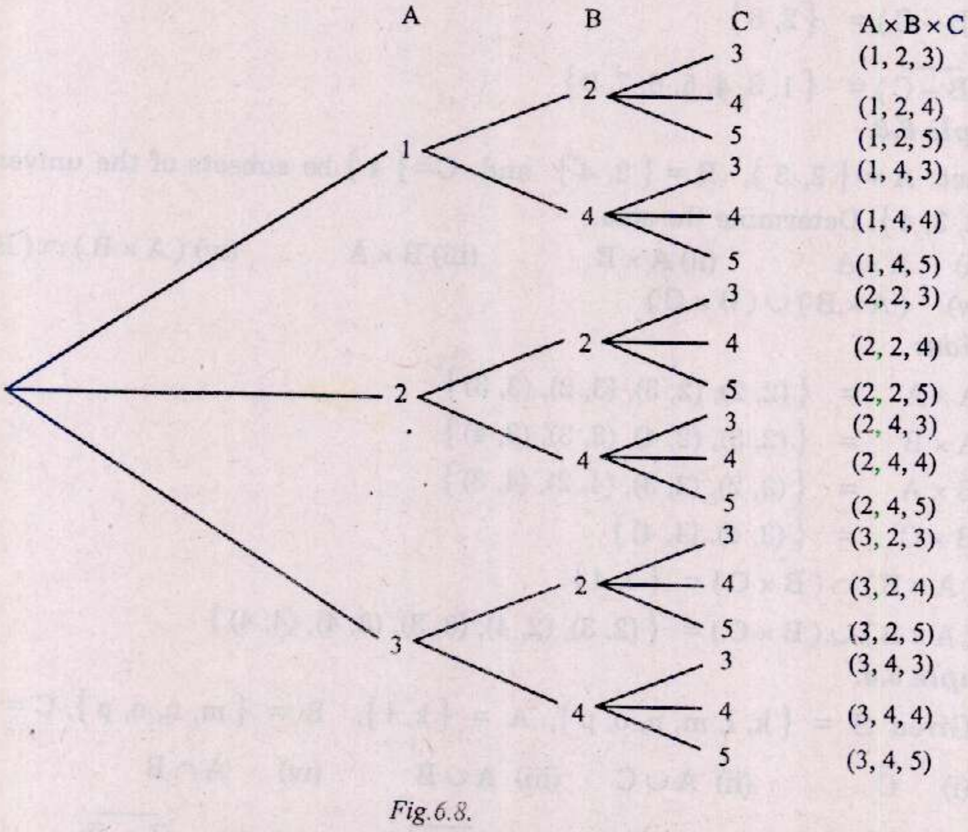


Fig.6.8.

6.9. ALGEBRA OF SETS

Having defined the operations of union, intersection, difference and complementation for sets, we may formulate an "algebra of sets". Many of the basic laws of set theory are analogous to the rules of algebra of real numbers.

(1) Identity Laws

(i) $A \cup \phi = A$ (ii) $A \cap \phi = \phi$ (iii) $A \cup S = S$ (iv) $A \cap S = A$

(2) Idempotent Laws

(i) $A \cup A = A$ (ii) $A \cap A = A$

(3) Associative Laws

(i) $(A \cup B) \cup C = A \cup (B \cup C)$

(ii) $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$

(4) Commutative Laws

(i) $A \cup B = B \cup A$ (ii) $A \cap B = B \cap A$

(5) Distributive Laws

(i) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (ii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(6) Complement Laws

(i) $A \cup \overline{A} = S$ (ii) $A \cap \overline{A} = \phi$

(iii) $\overline{(\overline{A})} = A$ (iv) $A \cap \phi = \phi$

(7) Demorgan's Laws

(i) $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$ (ii) $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$

6.10. MULTIPLICATION PRINCIPLE

The multiplication principle states that if an operation can be performed in n_1 ways, and then after it is performed in any one of these ways, a second operation can be performed in n_2 ways, and after this second operation is performed in any one of these ways, a third operation can be performed in n_3 ways, and so on for k operations, then the k operations can be performed in

$$n_1 \times n_2 \times n_3 \times \ldots \times n_k \text{ ways}$$

For example, in how many ways can the three letters A, B and C be arranged? By the principle of multiplication, we have

$$n_1 \times n_2 \times n_3 = 3 \times 2 \times 1 = 6 \text{ ways}$$

If a departmental store has five entrances and six exits, in how many ways can one enter and leave the store? Here, we write

$$n_1 \times n_2 = 5 \times 6 = 30 \text{ ways}$$

6.11. FACTORIALS

If n is a positive whole number, the product $n(n-1)(n-2) \ldots 3.2.1$ is called 'factorial n' and is denoted by the symbol n!. So

$$n! = n(n-1)(n-2) \ldots (3)(2)(1)$$

In particular, we see that

$$1! = 1$$
$$2! = 2 \times 1 = 2$$
$$3! = 3 \times 2 \times 1 = 6$$
$$4! = 4 \times 3 \times 2 \times 1 = 24$$
$$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3628800$$

Electronic calculators have the facility of calculating factorials. Zero factorial is equal to 1.

Thus $0! = 1$, but $(-4)!$, is not defined and has no meaning.

6.12. PERMUTATIONS

A permutation of a number of objects is an arrangement of these objects in a definite order. The number of permutations of a set of n things, taken r at a time is represented by ${}^nP_r$ and is given by

$${}^nP_r = \frac{n!}{(n-r)!} = n(n-1)(n-2) \ldots (n-r+1) \text{ ways}$$

For example, in how many different ways can the three letters a, b and c be arranged by taking two at a time? The answer is

$${}^3P_2 = \frac{3!}{(3-2)!} = 6 \text{ ways}$$

These permutations are ab, ba, ac, ca, bc, cb.

The number of permutations of a set of n things, taken all at the same time, is n!. Denoting this number by ${}^nP_n$, we have

$${}^nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

For example, in how many different ways can the 10 digits 0 to 9 be arranged by taking all of them at a time? The answer is

$${}^{10}P_{10} = 10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3628800$$

The number of permutations of n things consisting of groups among which n_1 are of one kind, n_2 of a second kind, n_3 of a third kind and so on, n_k of a kth kind is

$$\frac{n!}{n_1!\, n_2!\, n_3! \ldots n_k!} \quad \text{where } n_1 + n_2 + n_3 + \ldots + n_k = n$$

For example, in how many ways can the letters of the word "Samasatta" be arranged. Here, total number of letters is 9, 's' occurs 2 times, 'a' occurs 4 times, 't' occurs 2 times 'm' occurs 1 time. So the number of permutations is

$$\frac{9!}{2!\,4!\,2!\,1!} = \frac{9.8.7.6.5.4!}{2.1.4!\,2.1.1} = 3780 \text{ ways}$$

6.13. COMBINATIONS

A combination is a selection of objects considered without regard to their order. The total number of combinations of a set of n things taken r at a time, with $n \geq r$, usually denoted by ${}^nC_r$ or by $\binom{n}{r}$, is

$${}^nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!}$$

The number of combinations of the letters a, b and c taken two at a time is $\binom{3}{2} = 3$.

These combinations are ab, ac, bc.

Note: Here, ab is the same combination as ba.

A basket ball squad has 10 players. The coach must select a starting team. How many different teams of five players can be selected for this purpose? Here, we have

$${}^{10}C_5 = \binom{10}{5} = \frac{10!}{5!(10-5)!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5!}{5! \times 5 \times 4 \times 3 \times 2 \times 1} = 252 \text{ combinations}$$

The students are advised to remember the following relations

$${}^nC_1 = \binom{n}{1} = \frac{n!}{1!\,(n-1)!} = n$$

$${}^nC_n = \binom{n}{n} = \frac{n!}{n!\,(n-n)!} = \frac{n!}{n!\,0!} = 1$$

$${}^nC_{n-r} = \binom{n}{n-r} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

$$0! = 1$$

EXERCISES

Q.1 If $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, $A = \{0, 1, 3, 5, 7\}$,

$B = \{0, 2, 4, 6, 8\}$, $C = \{1, 2, 3, 4\}$ and $D = \{0, 5, 6\}$.

list the elements in the following sets.

(i) $A \cap B$ (ii) $C \cup A$ (iii) $\overline{D}$ (iv) $(C \cap \overline{D}) \cup B$,

(v) $\overline{(S \cap A)}$ (vi) $A \cap B \cap \overline{D}$.

Ans. (i) $\{0\}$ (ii) $\{0, 1, 2, 3, 4, 5, 7\}$ (iii) $\{1, 2, 3, 4, 7, 8\}$,

(iv) $\{0, 1, 2, 3, 4, 6, 8\}$ (v) $\{2, 4, 6, 8\}$ (vi) ϕ

Q.2 If $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{2, 4, 7, 9\}$, $B = \{1, 3, 5, 7, 9\}$, $C = \{2, 3, 4, 5\}$ and $D = \{1, 6, 7\}$. List the elements of the sets corresponding to the following events:

(i) $\overline{A} \cup C$ (ii) $B \cap \overline{C}$ (iii) $(\overline{S \cap \overline{B}})$ (iv) $(\overline{C} \cap D) \cup B$

(v) $(B \cap \overline{C}) \cup A$ (vi) $A \cap C \cap \overline{D}$

Ans. (i) $\{1, 2, 3, 4, 5, 6, 8\}$ (ii) $\{1, 7, 9\}$ (iii) $\{1, 3, 5, 7, 9\}$

(iv) $\{1, 3, 5, 6, 7, 9\}$ (v) $\{1, 2, 4, 7, 9\}$ (vi) $\{2, 4\}$

Q.3 Let $S = \{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, $A = \{4, 5, 6\}$, $B = \{5, 6, 7\}$ and $C = \{7, 8, 9\}$. List the members of the following events

(i) $\overline{A} \cap B$ (ii) $\overline{A} \cup B$ (iii) $(\overline{\overline{A} \cap \overline{B}})$ (iv) $\overline{A \cap (B \cup C)}$ (v) $\overline{S}$

Ans. (i) $\{7\}$ (ii) $\{3, 5, 6, 7, 8, 9, 10, 11, 12\}$ (iii) $\{4, 5, 6, 7\}$

(iv) $\{3, 4, 7, 8, 9, 10, 11, 12\}$ (v) $\{\ \} = \phi$

Q.4 Let $S = \{a, b, c, d, e, f, g, h, i, j\}$, $A = \{a, b, c, d\}$, $B = \{d, e, f\}$,

$C = \{a, c, e, i, j\}$. Find (i) $A \cap C$ (ii) $B \cup C$ (iii) $\overline{C} \cap \overline{A}$

(iv) $A \cap S$ (v) $\overline{(A \cup B)}$

Ans. (i) $\{a, c\}$ (ii) $\{a, c, d, e, f, i, j\}$ (iii) $\{f, g, h\}$ (iv) $\{a, b, c, d\}$

(v) $\{g, h, i, j\}$.

Q.5 Consider the sample space

$S = \{$ bus, train, aeroplane, boat, bicycle, automobile, motorcycle $\}$

and the events $A = \{$ bus, train, aeroplane $\}$,

$B = \{$ train, automobile, boat $\}$, $C = \{$ bicycle $\}$.

List the elements of the sets corresponding to the following events:

(i) $(\bar{A} \cup B)$ (ii) $(B \cap \bar{C}) \cap A$ (iii) $(\bar{A} \cup B) \cap (\bar{A} \cap \bar{C})$

Ans. (i) $\{$ train, boat, bicycle, automobile, motorcycle $\}$ (ii) $\{$ train $\}$

(iii) $\{$ boat, automobile, motorcycle $\}$

Q.6 If $S = \{x / 0 < x < 12\}$, $A = \{x / 1 < x < 9\}$ and $B = \{x / 0 < x < 5\}$, find:

(i) $A \cup B$ (ii) $A \cap B$ (iii) $\bar{A} \cap \bar{B}$

Ans. (i) $\{x / 0 < x < 9\}$ (ii) $\{x / 1 < x < 5\}$ (iii) $\{x / 8 < x < 12\}$

Q.7 Let $A = \{2, 3\}$, $B = \{1, 3, 5\}$ and $C = \{3, 4\}$. Construct the tree diagram of $A \times B \times C$ and then find $A \times B \times C$.

Ans. (2, 1, 3), (2, 1, 4), (2, 3, 3), (2, 3, 4), (2, 5, 3), (2, 5, 4), (3, 1, 3), (3, 1, 4), (3, 3, 3), (3, 3, 4), (3, 5, 3), (3, 5, 4).

Q.8 Let $S = \{a, b, c\}$, $T = \{b, c, d\}$ and $W = \{a, d\}$. Construct the tree diagram of $S \times T \times W$ and then find $S \times T \times W$.

Ans. (a, b, a), (a, b, d), (a, c, a), (a, c, d), (a, d, a), (a, d, d), (b, b, a), (b, b, d), (b, c, a), (b, c, d), (b, d, a), (b, d, d), (c, b, a), (c, b, d), (c, c, a), (c, c, d), (c, d, a), (c, d, d)

Q.9 Let $A = \{6, 7\}$, $B = \{7, 8\}$ and $C = \{8\}$ be subsets of the universal set $S = \{6, 7, 8\}$. Determine the elements of the following sets:

(i) $A \times C$ (ii) $A \times B$ (iii) $B \times A$ (iv) $(A \times B) \cup (A \times C)$

(v) $(A \times B) \cap (A \times C)$.

Ans. (i) $\{(6, 8), (7, 8)\}$ (ii) $\{(6, 7), (6, 8), (7, 7), (7, 8)\}$

(iii) $\{(7, 6), (7, 7), (8, 6), (8, 7)\}$ (iv) $\{(6, 7), (6, 8), (7, 7)\ (7, 8)\}$

(v) $\{(6, 8), (7, 8)\}$

Q.10 There are five different roads connecting two towns A and B. In how many ways can a person go from A to B by one road and return by another?

Ans. 20 ways

Q.11 How many lunches are possible consisting of soup, a sandwich, dessert and a drink if one can select from 5 soups, 2 sandwiches, 4 desserts and 3 drinks?

Ans. 120 ways

Q.12 There are five doors in a room. Three persons have to enter it. In how many ways can they enter from different doors?

Ans. 60 ways

Q.13 How many sample points are in the sample space when a pair of dice is thrown once?

Ans. 36 ways

Q.14 There are six seats available in a compartment. In how many ways can six persons be seated?

Ans. 720 ways

Q.15 Two lottery tickets are drawn from 25 for first and second prize. Find the number of sample points in the sample space S.

Ans. 600 sample points

Q.16 In how many different ways can a 11 – men football team be chosen from a squad of 20 men if the positions are not ignored.

Ans. 6.7044257×10^{12} ways

Q.17 How many possible permutations can be formed from the word Statistics?

Ans. 50400 ways

Q.18 In how many ways can the letters of the word "Mathematics" be arranged?

Ans. 4989600 ways

Q.19 How many possible permutations can be formed from the word Economics?

Ans. 90720 ways

Q.20 Evaluate (i) ${}^{9}P_4$ (ii) ${}^{7}P_5$ (iii) ${}^{16}P_2$ (iv) ${}^{5}P_5$ (v) ${}^{48}P_1$

Ans. (i) 3024 (ii) 2520 (iii) 240 (iv) 120 (v) 48

Q.21 Evaluate (i) $\binom{8}{4}$ (ii) $\dfrac{\binom{6}{3}\binom{4}{3}}{\binom{10}{6}}$ (iii) $\dfrac{\binom{7}{3}\binom{6}{2}}{\binom{13}{5}}$

(iv) $\dfrac{\binom{5}{3}}{\binom{13}{3}} \times \dfrac{\binom{8}{3}}{\binom{13}{3}}$ (v) $\dfrac{\binom{4}{4}\binom{48}{9}}{\binom{52}{13}}$ (vi) $\binom{8}{5}$

Ans. (i) 70 (ii) 0.3810 (iii) 0.4079 (iv) 0.0068 (v) 0.0026 (vi) 56

Q.22 In how many ways can a committee of 4 people be chosen out of 8 people?

Ans. 70

Q.23 A boy has five coins each of different denomination. How many different sums of money can he form?

Ans. 31

Chapter

7

PROBABILITY

7.1. INTRODUCTION

We live in a world of uncertainties. Man is surrounded by situations which are not fully under his control. The nature commands these situations. A person on a road does not know whether or not he will reach his destination safely. A patient in the hospital is never sure about his survival after a delicate operation. What will be the weather conditions tomorrow? A flight will be late or will reach in time, the road will be clear or there will be some traffic jam; we face such problems in our daily life. Man is always curious to know as to what will happen in future. The events which will happen in future are important for the man today. These events are based on what is called chance or probability. The numerical measure of uncertainty is called probability. We may find a numerical measure for a bulb to be defective, numerical measure for the rain to fall. The belief or confidence associated with a certain situation can also be measured. It is also called probability. In statistics, there are various situations where uncertainty is involved. Such situations need the application of probability. Probability is widely and rightly used in statistical decisions. The areas of statistics; where probability is used are called the areas of statistical inference. Statistical inference is not possible without the use of probability. Probability is used in all fields of life where uncertainty prevails. The knowledge of probability is used in space research, astronomy, business, weather forecast, economics, genetics and various other fields of life. It is simple to explain various concepts of probability with the help of set theory. Thus we shall use here the set theory notation which has been discussed in the previous chapter.

7.2. RANDOM EXPERIMENT

The word experiment or random experiment is used for a situation of uncertainty about which we want to have some observation. The actual result of the uncertain situation is called outcome or sample point. In the random experiment, nothing can be said with certainty about the outcome. An experiment may result in different outcomes, even though it is performed under similar conditions. The term random trial or simply trial is used, if an experiment is performed only once. A bulb may be selected from a factory to examine whether it is defective or not. The selection of a single bulb is a trial. We can select any number of bulbs. A random experiment has the following properties:

(i) The experiment can be repeated any number of times. We may select one or more than one items for inspection. The number of repetitions is called the size of the experiment. In statistics, the size of the random experiment plays a major role in statistical inference.

(ii) A random trial consists of at least two possible outcomes. If a basket contains all the defective bulbs, a selected bulb will be certainly defective. It has only one possible outcome. It is not a random trial. If the basket contains some good and some defective bulbs, a selected bulb will be good or defective. In this case there are two possible outcomes. Thus selecting a bulb from such a basket is a random trial.

(iii) Nothing can be said with certainty about the outcome of the random trial or random experiment. If a sample of four bulbs is selected, may be one bulb is defective. When another sample of four bulbs from the same lot is selected, may be all the bulbs are defective. Thus the result of the experiment cannot be predicted even if the experiment is repeated a large number of times.

7.3. SAMPLE SPACE

A complete list of all possible outcomes of a random experiment is called sample space or possibility space and is denoted by S. Each outcome is called element of the sample space. A sample space may contain any number of outcomes. If it contains finite number of outcomes, it is called finite or discrete sample space. When two bulbs are selected from a lot, the possible outcomes are four which can be counted as

(i) both bulbs are defective (ii) first is defective and second is good

(iii) first is good and second is defective. (iv) both are good.

Here the sample space is discrete. When the possibilities of the sample space cannot be counted, it is called continuous. The number of possible readings of temperature from 45°C to 46°C will make a continuous sample space.

Sample space is the basic term in the theory of probability. We shall discuss some sample spaces in this chapter. It is not always possible to make the sample space. If it contains very large number of points, we cannot register all the outcomes but we must understand as to how we can make the sample space. The outcomes of the sample space are written within the brackets { }. Some simple sample spaces are discussed below:

(i) A Coin is Tossed

When a coin is tossed, it has two possible outcomes. One is called head and the other is called tail. Anyone of the two faces may be called head. To be brief, head is denoted by H and tail is denoted by T. Thus the sample space consists of head and tail. In set theory notation, we can write S as:

$$S = \{\text{head, tail}\} \quad \text{or} \quad S = \{H, T\}.$$

(ii) Two Coins Tossed

When two coins are tossed, there are four possible outcomes. Let H_1 and T_1 denote the head and tail on the first coin and H_2 and T_2 denote the head and tail on

the second coin respectively. The sample space S can be written in the form of a table as below:

First coin	Second coin	
	H_2	T_2
H_1	(H_1H_2)	(H_1T_2)
T_1	(T_1H_2)	(T_1T_2)

This sample space can also be written as

$$S = \{(H_1H_2), (T_1H_2), (H_1T_2), (T_1T_2)\}$$

It may be noted here that a sample space of throwing two coins has 4 possible points. A sample space of 3 coins will have $2^3 = 8$ possible points and for n coins, the number of possible points will be 2^n.

(iii) A Die is Thrown

An ordinary die which is used in games of chances has six faces. These six faces contain 1, 2, 3, 4, 5, 6 dots on them. Thus for a single throw of a die, the sample space has 6 possible outcomes which are :

$$S = \{1, 2, 3, 4, 5, 6\}$$

(iv) Two Dice Thrown

A die has six faces. Each face of the first die can occur with all the six faces of the second die. Thus there are $6 \times 6 = 36$ possible pairs or points when two dice are thrown together. These 36 pairs are written below in Table 7.1.

Table 7.1.

First die	Second die					
	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

If 3 dice are thrown, the sample space will have $6^3 = 216$ possible outcomes where each outcome is composed of three digits.

7.3.1. EVENT OR SUBSET

Any part of the sample space is called an *event* or a *subset*. An *event* may contain one or more than one outcomes. When an *event* consists of a single outcome (sample point), it is called a *simple event*. An event which has two or more outcomes is called a *compound event*. The sample points contained in an event are written within brackets $\{\ \}$. If we consider a single face when a die is thrown, it is a *simple*

event. Getting 6 on a die when thrown, is called a simple event. If the event is any prime number on the die, the event consists of the points $\{ 2, 3, 5 \}$, which is a compound event and consists of three simple events which are $\{ 2 \}$, $\{ 3 \}$ and $\{ 5 \}$. When two dice are thrown, the pair (1, 1) is a single outcome in the sample space S and is therefore a *simple event*. The event "total is 3" consists of two outcomes that is (1, 2) and (2, 1). Thus "total is 3" is a compound event.

If a random experiment can produce n sample points, it has n simple events. Throw of a single die has 6 simple events and a throw of two dice produces 36 simple events.

The empty set ϕ is also an event but it is not a simple event.

The sample space S is a compound event and is called a certain event.

7.3.2. TOTAL NUMBER OF EVENTS

If a sample space has n sample points, then there are 2^n events which can be defined on this sample space. The number 2^n includes the empty set, n simple events and all possible compound events. Suppose two babies are born in a family. Let M denote the male baby and F denote the female baby. The sample space has four possible outcomes which are

$$S = \{ (M_1M_2), (M_1F_2), (F_1M_2), (F_1F_2) \}$$

The subscripts 1 and 2 stand for the first and second baby respectively.

All possible events which can be defined on this sample space are given in the following table.

Nature of Event	Outcomes	Number of events
Empty set	ϕ	1
Simple events	$\{M_1M_2\}, \{M_1F_2\}, \{F_1M_2\}, \{F_1F_2\}$	4
Compound events	$\{(M_1M_2), (M_1F_2)\}, \{(M_1M_2), (F_1M_2)\}$	
	$\{(M_1M_2), (F_1F_2)\}, \{(M_1F_2), (F_1M_2)\}$	6
	$\{(M_1F_2), (F_1F_2)\}, \{(F_1M_2), (F_1F_2)\}$	
	$\{(M_1M_2), (M_1F_2)\ (F_1M_2)\}$	
	$\{(M_1M_2), (M_1F_2), (F_1F_2)\}$	4
	$\{(M_1M_2), (F_1M_2), (F_1F_2)\}$	
	$\{(M_1F_2), (F_1M_2)\ (F_1F_2)\}$	
	$\{(M_1M_2), (M_1F_2), (F_1M_2), (F_1F_2)\}$	1
	Total possible events $= 2^n = 2^4 = 16$	

Example 7.1.

Two coins are tossed. Write all possible points of the sample space. Write the following events:

A_1 — both are heads A_2 — head on the first coin

A_3 — a head and a tail appears A_4 — at least one head appears

A_5 — both coins have the same faces A_6 — head on the first coin and tail on the second coin

Which of these events are simple?

Solution:

The sample space S is

$$S = \{(H_1H_2),\ (H_1T_2),\ (T_1H_2),\ (T_1T_2)\}$$

The outcomes in the events are

$$A_1 = \{(H_1H_2)\} \qquad A_2 = \{(H_1H_2), (H_1T_2)\}$$
$$A_3 = \{(H_1T_2), (T_1H_2)\} \qquad A_4 = \{(H_1H_2), (H_1T_2), (T_1H_2)\}$$
$$A_5 = \{(H_1H_2), (T_1T_2)\} \qquad A_6 = \{(H_1T_2)\}$$

The events A_1 and A_6 are simple and all other events are compound.

7.3.3. EQUALLY LIKELY OUTCOMES

The outcomes of a sample space are called *equally likely* if all of them have the same chance of occurrence. It is very difficult to decide whether or not the outcomes are equally likely. But in this book we shall assume in most of the experiments that the outcomes are equally likely. We shall apply the assumption of *equally likely* in the following cases.

(i) Tossing a Coin or Coins

When a coin is tossed, it has two possible outcomes called head and tail. We shall always assume that head and tail are equally likely if not otherwise mentioned. For more than one coin, it will be assumed that on all the coins, head and tail are equally likely.

(ii) Throwing a Die or Dice

Throw of a single die can produce six possible outcomes. All the six outcomes are assumed equally likely. For any number of dice, the six faces are assumed equally likely.

(iii) Playing Cards

There are 52 cards in a deck of ordinary playing cards. All the cards are of the same size and are therefore assumed equally likely.

(iv) Drawing Balls from a Bag

There are many situations in probability in which some objects are selected from a certain container. The objects of the container are assumed to be equally likely. A famous example is the selection of a few balls from a bag containing balls of different colours. The balls of the bag are assumed to be equally likely.

7.3.4. NOT EQUALLY LIKELY OUTCOMES

When all the outcomes of a sample space do not have equal chance of occurrence, the outcomes are called not equally likely. When a matchbox is thrown, all the six faces are not equally likely. If a bag contains balls of different sizes and a ball is selected at random, then all the balls are not equally likely.

7.3.5. MUTUALLY EXCLUSIVE EVENTS

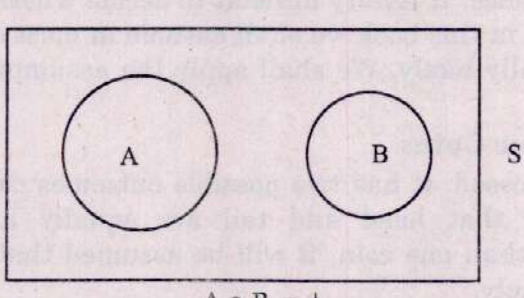
Two events are called mutually exclusive or disjoint if they do not have any outcome common between them. If the two events A and B are mutually exclusive, then $A \cap B = \phi$ (null set). For three mutually exclusive events A, B and C, we have $A \cap B \cap C = \phi$. Suppose there is a sample space S as:

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Let $A = \{\text{multiples of } 3\} = \{3, 6, 9\}$

$B = \{\text{multiples of } 5\} = \{5, 10\}$

Here $A \cap B = \phi$. Thus A and B are mutually exclusive events. Both A and B belong to the same sample space but they are completely different and both cannot happen at the same time. A class of students may contain first divisioners, second divisioners and third divisioners. When a student is selected from the class, he will be any one of the three groups of students. Thus three groups of students are disjoint or mutually exclusive. When the two events A and B are mutually exclusive, we can show them with the help of a venn diagram. The venn diagram in Figure 7.1. shows that $A \cap B = \phi$.



$A \cap B = \phi$

Fig. 7.1. A and B are Mutually Exclusive Events

7.3.6. NOT MUTUALLY EXCLUSIVE EVENTS

The events are called not mutually exclusive if they have at least one outcome common between them. If A and B are not mutually exclusive events, then $A \cap B \neq \phi$. Similarly A, B and C are not mutually exclusive events if $A \cap B \cap C \neq \phi$. Thus they must have at least one common point between them. Consider a sample space :

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$$

Let the event $A = \{\text{prime numbers}\} = \{2, 3, 5, 7, 11\}$

and $B = \{\text{odd faces}\} = \{1, 3, 5, 7, 9, 11\}$

Here $A \cap B = \{3, 5 \quad \}$

Thus $A \cap B \neq \phi$ i.e, $A \cap B$ exists. Here, A and B are not mutually exclusive events. $A \cap B$ consists of outcomes which are common to both A and B. Figure 7.2. shows a venn diagram in which A and B are not mutually exclusive events. Some area under A is common with B. If the event A is a part of the event B, then $A \cap B = A$. This is shown in Figure 7.3.

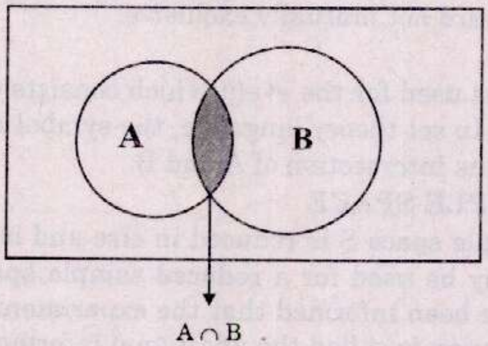
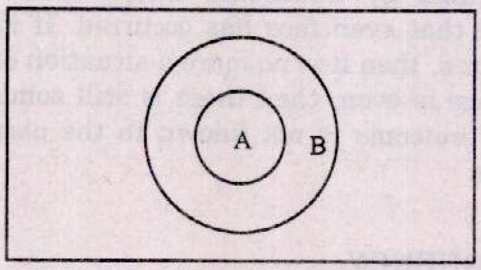


Fig. 7.2. A and B are Not Mutually Exclusive Events



$A \cap B = A$

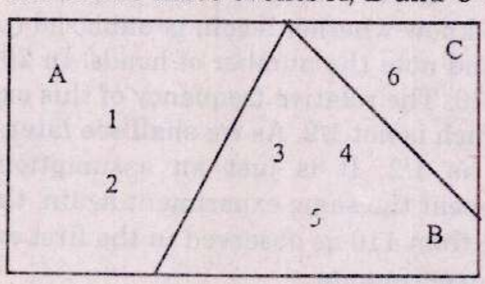
Fig.7.3. A and B are Not Mutually Exclusive Events

7.3.7. EXHAUSTIVE EVENTS

When a sample space S is partitioned into some mutually exclusive events such that their union is the sample space itself then the events are called *exhaustive events or collectively exhaustive events.* Suppose a die is tossed and the sample space is

$S = \{1, 2, 3, 4, 5, 6\}$. Let $A = \{1, 2\}$ $B = \{3, 4, 5\}$ $C = \{6\}$

Here the events A, B and C are mutually exclusive because $A \cap B \cap C = \phi$ and $A \cup B \cup C = S$. Figure 7.4. shows three events A, B and C which are exhaustive.



$A \cap B \cap C = \phi$ and $A \cup B \cup C = S$

Fig. 7.4.

7.3.8. A OR B

The term A or B is very important in probability theory. The term 'A or B' is used for two events A and B when they are mutually exclusive and we are interested to know the probability of the event "A or B". For this purpose the symbol $A \cup B$ is used. It must be remembered that the symbol $A \cup B$ is also used for the event "A or B or both", when A and B are not mutually exclusive.

7.3.9. A AND B

The term 'A and B' is used for the event which consists of the points which are common to both A and B. In set theory language, the symbol $A \cap B$ is used for A and B. The symbol $A \cap B$ means intersection of A and B.

7.3.10. REDUCED SAMPLE SPACE

Sometimes the sample space S is reduced in size and is called reduced sample space. The symbol S_r may be used for a reduced sample space. Suppose a die has been thrown and we have been informed that the experiment has produced an even face. This type of information is called the additional information. Thus the reduced sample space is determined by additional information. In this example the information has disclosed that even face has occurred. If it becomes known as to which even face has occurred, then it is no more a situation of probability. When the information is that the face is even, then there is still something hidden from the experimentor. The actual outcome is not known to the observer. In this case the reduced sample space S_r is

$$S_r = \{2, 4, 6\}$$

7.3.11. RELATIVE FREQUENCY

The term *relative frequency* is used for the ratio of the observed frequency of an outcome and the total frequency of the random experiment. Suppose a random experiment is repeated N times and an outcome is observed f times, then the ratio f / N is called the relative frequency of the outcome which has been observed f times. Some examples of relative frequencies are given here :

(i) We select bulbs from a certain big lot to examine whether they are good or defective. We take, say 100 such bulbs and examine them. Sixty bulbs are found defective. The symbol N may be used for 100 and the symbol f may be used for the observed frequency which is 60. Thus the

Relative Frequency $= f / N = 60/100 = 0.6$

(ii) We are interested to know whether a coin is unbiased (true) or not. We toss the coin say 200 times and note the number of heads. In 200 tosses, the number of heads may be, say 110. The relative frequency of this experiment for number of heads is 110/200 which is not 1/2. As we shall see later, the probability of head is usually written as 1/2. It is just an assumption and of course a big assumption. If we repeat the same experiment again, the number of heads may be less than or more than 110 as observed in the first experiment. This is what happens in random experiments.

(iii) A die is thrown and we are interested in the ace (face 1). We throw the die say 60 times and ace is observed 1[illegible] [illegible]nes. Thus the relative frequency of aces is

12/60 = 1/5. For an ideal die one should expect that the number of aces would be 60/6 = 10. At some later stage, we would like to know more about the ratio 12/60. This ratio is not something constant. A next random experiment with the same die may produce a completely different result.

(iv) The families having four children are being studied. Our study is based on 500 families. We have counted ,the families who do not have a male child. This observed frequency is 50. We have also counted the number of families having 1, 2, 3 and 4 male children, the number of families being 100, 200, 105 and 45. Thus the observed frequencies of families having 0, 1, 2, 3, 4 male children are 50, 100, 200, 105 and 45. This information and the relative frequencies can be written in the form of a table below:

Table 7.2.

Number of male children	Number of families (f)	Relative frequencies
0	50	$\frac{50}{500} = 0.10$
1	100	$\frac{100}{500} = 0.20$
2	200	$\frac{200}{500} = 0.40$
3	105	$\frac{105}{500} = 0.21$
4	45	$\frac{45}{500} = 0.09$
Total	500	1.00

In Table 7.2. we find the relative frequencies which are not equal. The students should not expect that the relative frequencies should be equal.

7.3.12. COMPLEMENTARY EVENTS

If A is an event defined in the sample space S, then S – A is denoted by $\bar{A}$ and is called complement of A.

Thus $\quad \bar{A} = S - A \quad$ or $\quad A \cup \bar{A} = S$

Figure 7.5. shows the event A and the complement of A.

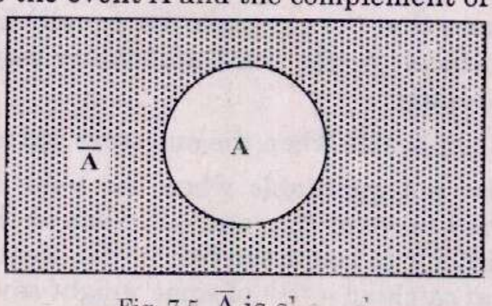


Fig. 7.5. $\bar{A}$ is s[illegible]

7.4. DEFINITIONS OF PROBABILITY

Probability is something strange and it has been defined in different manners. We can define probability in objective or subjective manner. Let us first use the objective approach to define probability.

7.4.1. THE CLASSICAL DEFINITION OF PROBABILITY

This definition is for equally likely outcomes. If an experiment can produce N mutually exclusive and equally likely outcomes out of which n outcomes are favourable to the occurrence of event A, then the probability of A is denoted by P(A) and is defined as the ratio n/N. Thus the probability of A is given by

$$P(A) = \frac{\text{Number of outcomes favourable to A}}{\text{Number of possible outcomes}} = \frac{n}{N}.$$

This definition can be applied in a situation in which all possible outcomes and the outcomes in the event A can be counted. This definition is due to P.S. Laplace (1749 – 1827). The classical definition is also called the priori definition of probability. The word priori is from prior and is used because the definition is based on the previous knowledge that the outcomes are *equally likely*. When a coin is tossed, the probability of head is assumed to be 1/2. This probability of 1/2 is based on this classical definition of probability. It is assumed that the two faces of the coin are equally likely. In practical life the people do believe that a coin will do justice when it is tossed. In the playgrounds, the participating teams toss the coin to start the match. A coin in which probability of head is assumed to be equal to the probability of tail is called a true or uniform or an unbiased coin. But it is all an assumption. The probability of a certain event is a number which lies between 0 and 1. If the event does not contain any outcome, it is called impossible event and its probability is zero. If the event is as big as the sample space, the probability of the event is 1 because

$$P(\text{the event}) = \frac{\text{Number of outcomes in the event}}{\text{Total number of outcomes}} = \frac{N}{N} = 1$$

When probability of an event is 1, it is called a 'sure' or 'certain', event.

CRITICISM

The *classical definition of probability* has always been criticised for the following reasons

(i) This definition assumes that the outcomes are equally likely. The term equally likely is almost as difficult as the word probability itself. Thus the definition uses the circular reasoning.

(ii) The definition is not applicable when the outcomes are not equally likely.

(iii) The definition is also not applicable when the total number of outcomes is infinite or it is difficult to count the total outcomes or the outcomes favourable to the event. It is difficult to count the fish in the ocean. Thus it is difficult to find the probability of catching a fish of some weight say more than 1 kg.

Example 7.2.

One day 20 files were presented to an income tax officer for disposal. Five files contained bogus entries. All the files were thoroughly mixed and there was no indication about bogus files. What is the probability that one file with bogus entries is selected.

Solution:

Here all possible outcomes = 20

Let A be the event that the file has bogus entries.

Thus, number of favourable outcomes = 5

Here we shall apply the classical definition of probability. All the 20 files are assumed to be equally likely for the purpose of selecting a file.

Probability of selecting a file with bogus entries is written as P(A)

$$\text{Thus, } P(A) = \frac{\text{Number of outcomes favourable to A}}{\text{Number of possible outcomes in S}} = \frac{n(A)}{n(S)} = \frac{5}{20} = \frac{1}{4}$$

Example 7.3.

A class of 60 students has 30 first divisioners, 20 second divisioners and 10 third divisioners. One student is selected at random. Find the probability that the selected student is (i) a first divisioner (ii) first or second divisioner (iii) second or third divisioner.

Solution:

There are sixty students in all. Thus there are 60 possible outcomes in S which are assumed to be equally likely. We shall apply the classical definition of probability to find the required probabilities

(i) Let A be the event that the selected student is a first divisioner. There are 30 outcomes favourable to the event A.

Thus the probability of the event A denoted by P(A) is

$$P(A) = \frac{\text{Number of outcomes favourable to A}}{\text{Number of possible outcomes in S}} = \frac{n(A)}{n(S)} = \frac{30}{60} = \frac{1}{2}$$

(ii) Let B be the event that the selected student is a first or second divisioner

$$\text{Hence, } P(B) = \frac{\text{Number of outcomes favourable to B}}{\text{Number of possible outcomes in S}} = \frac{n(B)}{n(S)} = \frac{50}{60} = \frac{5}{6}$$

(iii) Let C be the event that the selected student is a second or third divisioner. Then the number of outcomes favourable to C is 30.

$$\text{Hence, } P(C) = \frac{\text{Number of outcomes favourable to C}}{\text{Number of possible outcomes in S}} = \frac{n(C)}{n(S)} = \frac{30}{60} = \frac{1}{2}$$

Example 7.4.

A fair die is thrown. Find the probabilities that the face on the die is

(i) maximum (ii) prime (iii) multiple of 3 (iv) multiple of 7.

Solution:

There are 6 possible outcomes when a die is tossed. We assume that all the 6 faces are equally likely. The classical definition of probability is to be applied here.

The sample space is

$$S = \{1, 2, 3, 4, 5, 6\}, \quad n(S) = 6$$

(i) Let A be the event that the face is maximum

Thus, $A = \{6\}$, $n(A) = 1$, $\quad P(A) = \dfrac{n(A)}{n(S)} = \dfrac{1}{6}$

(ii) Let B be the event that the face is prime, then

$$B = \{2, 3, 5\}, \quad n(B) = 3, \quad P(B) = \frac{n(B)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

(iii) Let C be the event that the face is multiple of 3

Hence, $C = \{3, 6\}$, $n(C) = 2$, $\quad P(C) = \dfrac{n(C)}{n(S)} = \dfrac{2}{6} = \dfrac{1}{3}$

(iv) Let D be the event that face is multiple of 7. There is no multiple of 7 in the sample space. Thus $D = \phi$, $P(D) = \dfrac{n(D)}{n(S)} = 0$ (not possible)

Example 7.5.

Two dice are rolled. Let A be the event that both faces are same, B be the event that total on the two dice is less than 5 and C be the event that there is at least one ace on the two dice. Write the sample space S and the outcomes which belong to A, B, C, $A \cup B$, $A \cap B$ and find their probabilities.

Solution:

When two dice are rolled, there are 36 possible outcomes which are given in Table 7.1.

The events (sub-sets) A, B, C, $A \cup B$ and $A \cap B$ are written as:

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}, \; n(A) = 6$$

$$B = \{(1, 1), (1, 2), (2, 1), (1, 3), (3, 1), (2, 2)\}, \; n(B) = 6$$

$$C = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1)\}, \; n(C) = 11$$

$$A \cup B = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 2), (2, 1), (1, 3), (3, 1)\}, \; n(A \cup B) = 10$$

$$A \cap B = \{(1, 1), (2, 2)\}, \; n(A \cap B) = 2$$

Hence $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{6}{36} = \dfrac{1}{6}$

$$P(B) = \frac{n(B)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{11}{36}$$

$$P(A \cup B) = \frac{\text{Number of outcomes in } A \cup B}{\text{Number of all possible outcomes}} = \frac{n(A \cup B)}{n(S)} = \frac{10}{36} = \frac{5}{18}$$

$$P(A \cap B) = \frac{\text{Number of outcomes in } A \cap B}{\text{Number of all possible outcomes}} = \frac{n(A \cap B)}{n(S)} = \frac{2}{36} = \frac{1}{18}$$

Note: The events $(A \cup B)$ and $(A \cap B)$ are the events which belong to S and we can find their probabilities by using the classical definition of probability.

Example 7.6.

Suppose a fair die is rolled. Let A be the event that the face on the die is prime and B be the event that the face is even. Write the elements of the following events and find their probabilities.

(i) A (ii) B (iii) $\bar{A}$ (iv) $\bar{B}$ (v) $A \cap B$

(vi) $A \cup B$ (vii) $\overline{A \cap B}$ (viii) $\overline{A \cup B}$ (ix) $\overline{\overline{A \cup B}}$

Solution:

When a die is rolled, the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

As the die is fair, therefore all the 6 outcomes are equally likely. Each outcome has probability of 1/6. We shall apply the classical definition of probability to find the probabilities of the given events.

Thus $A = \{2, 3, 5\}$ $B = \{2, 4, 6\}$

$\bar{A} = \{1, 4, 6\}$ $\bar{B} = \{1, 3, 5\}$

$A \cap B = \{2\}$ $A \cup B = \{2, 3, 4, 5, 6\}$

$$\overline{A \cap B} = S - A \cap B = \{1, 3, 4, 5, 6\}$$

$$\overline{A \cup B} = S - A \cup B = \{1\}$$

$$\overline{\overline{A \cup B}} = S - \overline{A \cup B} = \{2, 3, 4, 5, 6\} = A \cup B$$

Hence, (i) $P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$ (ii) $P(B) = \frac{n(B)}{n(S)} = \frac{3}{6} = \frac{1}{2}$

(iii) $P(\bar{A}) = \frac{n(\bar{A})}{n(S)} = \frac{3}{6} = \frac{1}{2}$ (iv) $P(\bar{B}) = \frac{n(\bar{B})}{n(S)} = \frac{3}{6} = \frac{1}{2}$

(v) $P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{6}$ (vi) $P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{5}{6}$

(vii) $P(\overline{A \cap B}) = \frac{n(\overline{A \cap B})}{n(S)} = \frac{5}{6}$ (viii) $P(\overline{A \cup B}) = \frac{n(\overline{A \cup B})}{n(S)} = \frac{1}{6}$

(ix) $P(\overline{\overline{A \cup B}}) = \frac{n(\overline{\overline{A \cup B}})}{n(S)} = \frac{5}{6} = P(A \cup B)$

Example 7.7.

A bag contains 5 white, 4 black, 3 green and 6 red balls of the same size and weight. A ball is selected at random, find the following probabilities.

(i) The selected ball is white. (ii) The selected ball is white or black.

(iii) The ball is not black.

Solution:

As all the balls are of the same size, the 18 possible outcomes are equally likely. A probability of 1/18 is attached to each ball. The classical definition of probability is applied here.

Thus, (i) P(ball is white) $= \frac{5}{18}$ (ii) P(ball is white or black) $= \frac{9}{18} = \frac{1}{2}$

(iii) P(ball is not black) $= \frac{14}{18} = \frac{7}{9}$.

Example 7.8.

Two uniform coins are tossed. Find the following probabilities:

(i) both are heads (ii) both faces are same (iii) only one is head

(iv) at least one is head (v) head on the first coin.

Solution:

When two coins are tossed, there are four possible outcomes. Thus the sample space is

$$S = \{(HH), (TH), (HT), (TT)\}, \; n(S) = 4$$

All the 4 outcomes are equally likely. Each outcome has probability of 1/4. The outcome (TH) has tail T of the first coin and head H of the second coin.

(i) Let A denote the event that both coins give heads, then

$$A = \{(HH)\}, n(A) = 1, \; P(A) = \frac{n(A)}{n(S)} = \frac{1}{4}$$

(ii) Let B be the event that both faces are same, then

$$B = \{(HH), (TT)\}, n(B) = 2, \; P(B) = \frac{n(B)}{n(S)} = \frac{2}{4} = \frac{1}{2}$$

(iii) Let C denote the event that only one is head, then

$$C = \{(HT), (TH)\}, n(C) = 2, \; P(C) = \frac{n(C)}{n(S)} = \frac{2}{4} = \frac{1}{2}$$

This event is also called "just one is head" or "exactly one is head". We can also say "only one is tail" or "just one is tail". All these statements mean the same event C.

(iv) Let D denote the event that at least one is head. It means any number of heads, thus

$$D = \{(HH), (HT), (TH)\}, n(D) = 3, \; P(D) = \frac{n(D)}{n(S)} = \frac{3}{4}$$

(v) Let E stand for the event "head on the first coin". The event E means that anything may come on the second coin but it should be head on the first coin. Thus

$$E = \{(HH), (HT)\}, n(E) = 2, \ P(E) = \frac{n(E)}{n(S)} = \frac{2}{4} = \frac{1}{2}$$

Example 7.9

A card is selected from a deck of playing cards. Find the following probabilities:

(i) the card is red (ii) the card is a king (iii) the card is of diamonds

(iv) the card is a king or queen (v) it is a faced card.

Solution:

A deck of playing cards has 52 cards which are of the same size. When a card is selected, there are 52 possible outcomes which are assumed to be equally likely. Thus the classical definition of probability is applied here to find the probabilities of the events belonging to this sample space.

Here, total number of possible outcomes $= n(S) = 52$

(i) The deck has 26 red cards, therefore

$$P(\text{card is red}) = \frac{\text{Number of red cards}}{\text{Total number of cards}} = \frac{26}{52} = \frac{1}{2}$$

(ii) There are 4 kings in the playing cards, therefore

$$P(\text{card is a king}) = \frac{\text{Number of kings}}{\text{Total number of cards}} = \frac{4}{52} = \frac{1}{13}$$

(iii) There are 13 cards of diamonds, thus the event contains 13 points out of 52 points all of which are equally likely. Therefore

$$P(\text{card is of diamonds}) = \frac{13}{52} = \frac{1}{4}$$

(iv) There are 4 kings and 4 queens, thus the event contains 8 points. Hence,

$$P(\text{card is a king or queen}) = \frac{8}{52} = \frac{2}{13}$$

(v) The jacks, queens and kings are called faced or pictured cards. Thus there are 12 faced cards. Hence,

$$P(\text{card is faced}) = \frac{12}{52} = \frac{3}{13}$$

Example 7.10.

A digit is selected at random from the first ten natural numbers. Find the probability that the selected digit is (i) greater than 6 (ii) a complete square (iii) multiple of 3 (iv) a prime number less than 3.

Solution:

There are 10 possible outcomes in the sample space S. It is assumed here that all the 10 outcomes of the sample space are equally likely. Each digit has probability

of selection equal to 1/10. Thus we shall apply here the classical definition of probability.

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Total number of possible outcomes $= 10 = n(S)$

(i) Subset of digits greater than 6 $= \{7, 8, 9, 10\}$

$$P(\text{ digit is greater than 6 }) = \frac{\text{Number of digits in the subset}}{\text{Total number of digits}} = \frac{4}{10} = \frac{2}{5}.$$

(ii) Subset of digits which are a complete square $= \{1, 4, 9\}$

$$P(\text{ digit is a complete square }) = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}} = \frac{3}{10}$$

(iii) Subset of multiples of 3 $= \{3, 6, 9\}$

$$P(\text{ digit is multiple of 3 }) = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}} = \frac{3}{10}$$

(iv) Subset of a prime number less than 3 $= \{2\}$

P(digit is a prime number less than 3)

$$= \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}} = \frac{1}{10}$$

7.4.2 THE RELATIVE FREQUENCY DEFINITION

This definition is based on the actual observations when the random experiment is repeated a large number of times under similar conditions. Suppose a random experiment is repeated N times and it is observed that a certain event say A has actually occurred n times, then the probability of the event A is defined as the ratio n / N where N is very large.

In symbols, we write $P(A) = \lim_{N \to \infty} \frac{n}{N}$

This definition is also called the *empirical* or the statistical definition because we observe the occurrence of the event when the experiment is being repeated. It is assumed in this definition that the ratio n/N will become a stable value when N is repeated a large number of times. This stable value of the ratio is called the probability of A and is written as P(A).

This definition is closer to the real world situations. If we are interested to know whether or not a coin is true, we can toss a coin a large number of times and the number of heads can be counted. The probability of head is given by

$$P(\text{ head }) = \frac{\text{Number of heads actually observed}}{\text{Total number of throws}}$$

This ratio may or may not be equal to 1/2. It is not possible to find this ratio by classical definition. Suppose we take a sample of 100 bulbs from a certain factory and examine them. We find that 20 are defective. The ratio 20/100 is called the proportion of defective bulbs. This ratio is 0.2. It is also called the probability of

selecting a defective bulb provided sample size is increased and the ratio reaches some stable value. This definition can be applied in real world problems for finding the probabilities. But this definition also has certain deficiencies such as :

(i) The experiment is to be repeated a large number of times under the same conditions. It is difficult to maintain exactly the same conditions for lengthy experiments. When a coin is tossed a large number of times, the coin may reduce in size. The surface on which the experiment is done may not remain the same.

(ii) Repeating the experiment a very large number of times is a very time consuming and sometimes very costly affair. Suppose we want to find the probability that a bullet or a missile will fire. It is not possible to repeat the experiment an infinite number of times because a lot of expenditure is involved on this experiment.

Example 7.11.

Mr. A has played 10 games with Mr. B in the past. Six games were won by Mr. A. They have decided to play a game. What is the probability that Mr. A will win the game.

Solution:

An experiment has been conducted in the past and it has been observed that the event (Mr. A wins) has occurred 6 times. The experiment has been repeated 10 times and the event has occurred 6 times. We apply here the relative frequency definition of probability to find the probability that the game they are going to play will be won by Mr. A.

Thus, P(A will win the game) = $6/10 = 0.6$

Example 7.12.

A coin has been tossed different number of times and number of heads have been recorded. The results of the experiment are given below. What do you think is the probability of getting head on the coin.

Number of throws	Number of heads
50	30
100	70
200	125
300	180
⋮	⋮
5000	2560
10000	5100
20000	10202

Solution:

The random experiment has been conducted to find the probability of head on a coin. We find the relative frequencies of heads in the following table.

Number of throws (N)	Number of heads (f)	Relative frequency of heads $\left(\frac{f}{N}\right)$
50	30	$\frac{30}{50} = 0.6000$
100	70	$\frac{70}{100} = 0.7000$
200	125	$\frac{125}{200} = 0.6250$
300	180	$\frac{180}{300} = 0.6000$
$\vdots$	$\vdots$	$\vdots$
5000	2560	$\frac{2560}{5000} = 0.5120$
10000	5100	$\frac{5100}{10000} = 0.5100$
20000	10202	$\frac{10202}{20000} = 0.5101$

According to the relative frequency definition of probability, the ratio f / N is called the probability of the event provided the ratio f / N reaches some stable value. In this experiment we see that the ratio f / N becomes stable at 0.51 when the coin is tossed more than 5000 times. Thus we can say that the probability of head is 0.51 for the coin with which the experiment has been conducted. When the probability of head is not 1/2, the coin is called biased.

7.4.3. THE AXIOMATIC DEFINITION OF PROBABILITY

This definition is based on certain axioms which must be satisfied by the probability of an event. Suppose a sample space S has the sample points $E_1, E_2, \ldots, E_i, \ldots, E_n$. Each point is assigned a real number like $P(E_1), P(E_2), \ldots, P(E_i), \ldots, P(E_n)$ which are called the weights or probabilities of the points. Probability of any point denoted by $P(E_i)$ must obey the following axioms :

(i) $0 \le P(E_i) \le 1$

The probability of the event must lie between 0 and 1.

(ii) $P(S) = 1$ for the event which is as large as the sample space.

(iii) If A and B are mutually exclusive events, then

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

According to this definition, a non-negative number is assigned to each point in S such that sum of all such numbers is 1. In this definition the condition of equally likely outcomes is not necessary. The outcomes may or may not be equally likely. When they are equally likely, each point in S will have the same probability.

Probability of an Event A

If an event A belongs to the sample space S, then the probability of A denoted by $P(A)$ is the sum of the probabilities of all sample points E_i included in A.

$$\text{Thus } P(A) = \Sigma P(E_i)$$

It is possible that all E_i are equally likely. In this case $P(E_1) = P(E_2) = \ldots = P(E_n)$. Each probability is 1/n because these are n equally likely points.

If E_1, E_2, E_3 are included in the event A, then

$$P(A) = P(E_1) + P(E_2) + P(E_3) = \frac{1}{n} + \frac{1}{n} + \frac{1}{n}$$

$$= \frac{3}{n} = \frac{\text{Number of sample points in A}}{\text{Number of sample points in S}}$$

Example 7.13.

A coin has been designed in such a manner that the head has been assigned a weight of 3w and tail has been assigned a weight of 2w. Find the probability of head.

Solution:

According to the axiomatic definition of probability, the total weight of all possible outcomes must be 1. We can write

Outcomes	Weight
Head	3 w
Tail	2 w

Total weight $= 3w + 2w = 1$, $5w = 1$ or $w = 1/5$

Hence P(head) $= 3w = 3 \times 1/5 = 3/5 = 0.6$

and P(tail) $= 2w = 2 \times 1/5 = 2/5 = 0.4$

We can say that the coin is not true.

Example 7.14.

Six faces of an unbalanced die are assigned weights as below. Find probability of each face.

Face	1	2	3	4	5	6
Weight	w	2w	w	2w	w	2w

Solution:

We shall apply here the axiomatic approach to probability. The total weight of all possible outcomes must be 1. Thus

Total weight $= w + 2w + w + 2w + w + 2w = 1$ or $9w = 1$ and $w = 1/9$

Thus the six faces of the die have the probabilities as below:

Face	1	2	3	4	5	6
Probability	1/9	2/9	1/9	2/9	1/9	2/9

The probability of each odd face is 1/9 and the probability of each even face is 2/9.

Note: According to the classical definition of probability, the probability of head on a coin is always 1/2. In Example 7.12, the probability of head has been calculated by relative frequency definition and it is 0.51 for a certain coin. This calculation is for the particular coin which has been tossed 20000 times. In Example 7.13, the probability of head has been calculated by axiomatic approach and it is 0.6 for a certain coin.

7.4.4. SUBJECTIVE OR PERSONALISTIC PROBABILITY

A person may have some confidence or belief regarding the occurrence of some event, say A. The numerical measure of this confidence is called the subjective probability of the occurrence of A. This probability is based on the experience, intelligence and knowledge of the person who is determining the probability in some situation. For example, we may be interested to know whether a certain political system will succeed in a country or not. The probability of success in this situation cannot be determined by objective definitions of probability. The assessment of this probability is made by some expert. This approach can be applied in real world situations. This probability is subjective in nature. Different persons may have different probabilities for the same situation at the same time.

7.5. ADDITION LAW FOR MUTUALLY EXCLUSIVE EVENTS

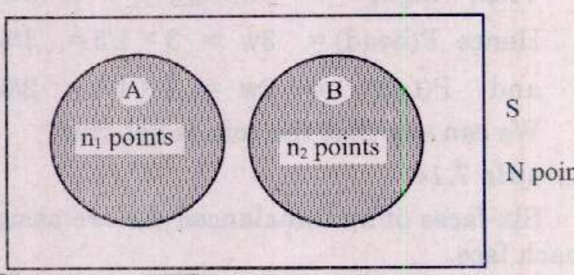
Theorem:

If the two events A and B are mutually exclusive then the probability that any one of them will occur is the sum of the probabilities of A and B. In symbols,

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

Proof:

Suppose there is a sample space S containing N points (outcomes) which are equally likely. The two events A and B belong to S and contain n_1 and n_2 points respectively. There is no point common to A and B. As A and B are disjoint, the union of A and B contains $(n_1 + n_2)$ points. Using classical definition of probability, we can write



$A \cup B$ is shaded

Fig.7.7.

$$P(A \cup B) = \frac{\text{Number of outcomes favourable to } A \cup B}{\text{Total number of outcomes in S}}$$

$$= \frac{n_1 + n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(A) + P(B)$$

Thus $P(A \cup B) = P(A) + P(B)$

The law can be generalised for more than two events. If $A_1, A_2, \ldots, A_k$ are k mutually exclusive events then

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k) = \sum_{i=1}^{k} P(A_i)$$

As a special case, if $A_1 \cup A_2 \cup \ldots \cup A_r = S$

then $P(A_1 \cup A_2 \cup \ldots \cup A_r) = P(S) = 1$

$P(A_1) + P(A_2) + \ldots + P(A_r) = 1$

Here $A_1, A_2, \ldots, A_r$ are called exhaustive events. They are mutually exclusive and they cover the whole sample space.

Corollary 1: If ϕ is an empty set, then $P(\phi) = 0$

We know $\phi \cup S = S$ and $P(\phi \cup S) = P(S)$

ϕ and S are mutually exclusive events, therefore

$P(\phi) + P(S) = P(S)$ or $P(\phi) = P(S) - P(S) = 0$

Corollary 2: If $\bar{A}$ is complement of a set A relative to same sample space S, then

$A \cup \bar{A} = S$

and $P(A \cup \bar{A}) = P(S) = 1$

A and $\bar{A}$ are mutually exclusive, therefore

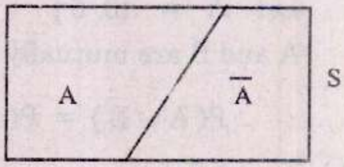


Fig. 7.8.

$P(A) + P(\bar{A}) = 1$ or $P(\bar{A}) = 1 - P(A)$

This is called law of complementation of two sets.

Example 7.15.

A coin is tossed 6 times. Find the probability that there is at least one head.

Solution:

The sample space S has $2^6 = 64$ possible outcomes. There is only one outcome (TTTTTT) in S which does not have any head. All the other 63 outcomes have either one or more heads. It is difficult to write all the 63 outcomes. Let A be the event "at least one head" and $\bar{A}$ be the event "no head".

A and $\bar{A}$ are mutually exclusive and exhaustive events, therefore

$A \cup \bar{A} = S$ and $P(A \cup \bar{A}) = P(S)$

$P(A) + P(\bar{A}) = 1$ or $P(A) = 1 - P(\bar{A}) = 1 - 1/64 = 63/64$

Hence, P(at least one head) = 63/64

Example 7.16.

A card is selected at random from a deck of playing cards. Find the probability that the card is a king or a queen.

Solution:

There are 52 possible outcomes which are equally likely. Each card has probability of 1/52.

Let A represent the event "card is a king" and B represent the event "card is a queen". There are 4 kings and 4 queens in a deck. The events A and B are mutually exclusive, therefore

$$P(A \cup B) = P(A) + P(B) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52}$$

We have already solved this question in *Example 7.9* but here we have used a law called the addition law of disjoint events.

Example 7.17.

A die is rolled. Find the following probabilities.

(i) The face is multiple of 3 or multiple of 5.

(ii) The face is a complete square or it is the maximum face.

Solution:

When a die is rolled, the sample space, is $S = \{1, 2, 3, 4, 5, 6\}$

(i) Let $A = \{3, 6\}$ $B = \{5\}$

A and B are mutually exclusive events, therefore

$$P(A \cup B) = P(A) + P(B) = \frac{2}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Alternately

$$A = \{3, 6\} \qquad B = \{5\} \qquad A \cup B = \{3, 5, 6\}$$

From the classical definition of probability of an event, we have

$$P(A \cup B) = \frac{\text{Number of outcomes in } A \cup B}{\text{Total number of possible outcomes}} = \frac{n(A \cup B)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

(ii) Let $A = \{1, 4\}$ $B = \{6\}$

A and B are mutually exclusive events, therefore

$$P(A \cup B) = P(A) + P(B) = \frac{2}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Alternately

$$A = \{1, 4\} \qquad B = \{6\} \qquad A \cup B = \{1, 4, 6\}$$

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

When the events A and B contain small number of points, we can list the points in A, B and $A \cup B$ and we can find the probability of the event $A \cup B$ by using the definition of probability i.e.,

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)}$$

But when it is difficult to list the elements in $A \cup B$, we would use the addition law of probability for mutually exclusive events.

Example 7.18.

Two dice are rolled. Find the probability that the total is less than 12.

Solution:

When two dice are rolled, there are 36 possible points. In the sample space there are 35 (please see Table 7.1) pairs in which the total is less than 12. Let A represent these 35 outcomes and $\bar{A}$ represents only one outcome (the total of 12)

$$P(\bar{A}) = 1/36$$

As A and $\bar{A}$ are disjoint. Therefore $P(A \cup \bar{A}) = P(S) = 1$

$$P(A) + P(\bar{A}) = 1 \quad \text{or} \quad P(A) = 1 - P(\bar{A}) = 1 - 1/36 = 35/36$$

Example 7.19.

If A and B are two mutually exclusive events from a sample space, then is it possible that $P(A) = 0.7$ and $P(B) = 0.6$?

Solution:

As A and B are mutually exclusive, therefore

$A \cup B \leq S$

If A and B are exhaustive events,

then $A \cup B = S$

If A and B are not exhaustive events,

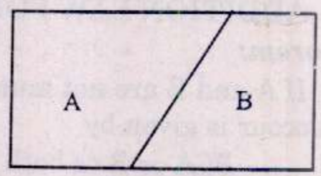
then $A \cup B < S$

Thus $A \cup B \leq S$

and $P(A \cup B) \leq P(S)$

$P(A) + P(B) \leq 1$

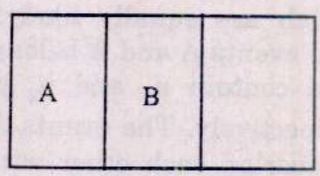
$0.7 + 0.6 \nleq 1$



$A \cup B = S$

A and B are exhaustive events

Fig.7.9.



$A \cup B < S$

A and B are not exhaustive events

Fig.7.10.

Hence A and B are not mutually exclusive events. The given probabilities are wrong.

Example 7.20.

A digit is selected at random from the first 100 natural numbers. Find the probability that the selected digit is multiple of 10 or multiple of 11.

Solution:

Let A be the event "digit is multiple of 10" and B be the event "digit is multiple of 11".

Then, $n(S) = 100$

$n(A) = 100/10 = 10$ points (10, 20, 30, 40, 50, 60, 70, 80, 90, 100)

$n(B) = 100/11 = 9$ points (11, 22, 33, 44, 55, 66, 77, 88, 99)

Any multiple of 10 is not common with the multiples of 11. Thus A and B are mutually exclusive, therefore

$$P(A \cup B) = P(A) + P(B) = \frac{10}{100} + \frac{9}{100} = \frac{19}{100}$$

Example 7.21

A, B and C are taking part in a race. The chance of winning the race by A is half of that of B and B winning the race is half of that of C. Find their respective chances of winning the race.

Solution:

Let the probability of winning the race by A is p i.e;

$P(A) = p$, then $P(B) = 2p$ and $P(C) = 4p$.

Since A, B and C are mutually exclusive events, therefore

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

$$1 = p + 2p + 4p \quad \text{or} \quad 7p = 1 \quad \text{or} \quad p = 1/7$$

Hence $P(A) = 1/7$, $P(B) = 2/7$ and $P(C) = 4/7$

7.6 ADDITION LAW FOR NOT MUTUALLY EXCLUSIVE EVENTS

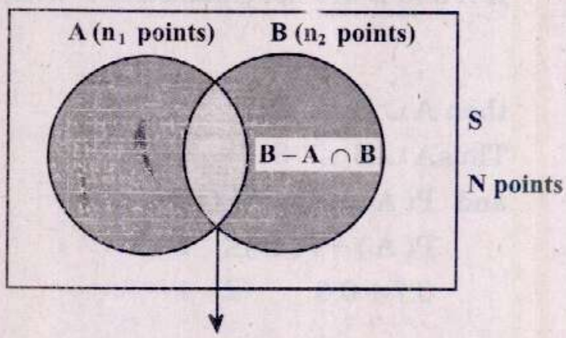
Theorem:

If A and B are not mutually exclusive events then the probability that A or B or both occur is given by

$$P(A \text{ or } B \text{ or both}) = P(A \cup B) = P(A) + P(B) - P(A \cup B)$$

Proof:

Suppose there is a sample space S containing N points which are equally likely. The two events A and B belong to S and contain n_1 and n_2 points respectively. The events A and B overlap each other with the result that $A \cap B$ has m points. Thus m points are common to A and B. $A \cup B$ can be written as union of two disjoint events which are



$A \cap B$ has m points

$A \cup B$ is shaded

Fig.7.11.

(i) A (ii) $(B - A \cap B)$

Thus we can write $A \cup B = A \cup (B - A \cap B)$

A contains n_1 points and $(B - A \cap B)$ contains $(n_2 - m)$ points

Thus $A \cap B$ has $n_1 + (n_2 - m)$ points

Using classical definition of probability, we have

$$P(A \cup B) = \frac{\text{Number of points in } A \cup B}{\text{Number of points in } S} = \frac{n_1 + n_2 - m}{N} = \frac{n_1}{N} + \frac{n_2}{N} - \frac{m}{N}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

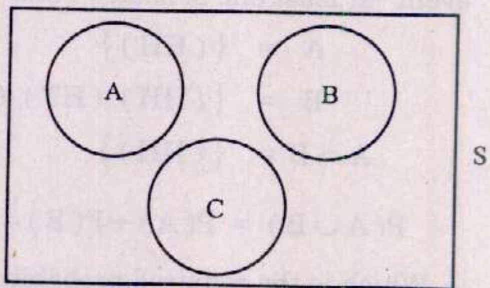
Which is called the general law of addition of probabilities.

When A and B are mutually exclusive, then $A \cap B = \phi$ and $P(A \cap B) = 0$, we get $P(A \cup B) = P(A) + P(B)$ which is the addition law for mutually exclusive events. For three not mutually exclusive events,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

When A, B and C are mutually exclusive, then

$$\begin{aligned} P(A \cap B) &= P(A \cap C) \\ &= P(B \cap C) \\ &= P(A \cap B \cap C) \\ &= 0 \end{aligned}$$



$A \cap B \cap C = \phi$

A, B, C are mutually exclusive

Fig.7.12.

Therefore, $P(A \cap B \cap C) = P(A) + P(B) + P(C)$

This is called addition law of probability for mutually exclusive events.

Example 7.22.

A die is tossed. Find the probability that the face is a prime or is even number.

Solution:

When a die is tossed, there are 6 possible outcomes which are equally likely. Thus

$$S = \{1, 2, 3, 4, 5, 6\} \text{ and } n(S) = 6$$

Let A represent the event "face is prime" and B represent the event "face is even" then

$$A = \{2, 3, 5\} \quad B = \{2, 4, 6\} \quad A \cap B = \{2\}$$

A and B are overlapping. Using the addition law for not mutually exclusive events we have, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$

Alternate Solution

We have already solved this question in Example 7.6.

We have $S = \{1, 2, 3, 4, 5, 6\}$ $A = \{2, 3, 5\}$ $B = \{2, 4, 6\}$

$$A \cup B = \{2, 3, 4, 5, 6\} \qquad P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{5}{6}$$

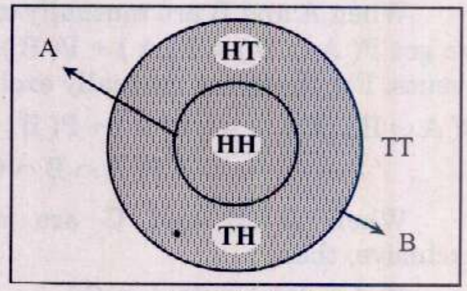
Example 7.23.

Two coins are tossed. Find the probability that both faces are heads or at least one is head.

Solution:

When two coins are tossed, the sample space S contains $2^2 = 4$ possible outcomes. Let A represent the event "both are heads" and B represent the event "at least one is head". Thus

$$A = \{(HH)\}$$
$$B = \{(HH), (HT), (TH)\}$$
$$A \cap B = \{(HH)\}$$



A ∪ B is shaded

Fig.7.13.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{4} + \frac{3}{4} - \frac{1}{4} = \frac{3}{4}$$

Which is the required probability.

Alternately

We have $S = \{(HH), (HT), (TH), (TT)\}$ $A = \{(HH)\}$

$B = \{(HH), (HT), (TH)\}$ $A \cup B = \{(HH), (HT), (TH)\}$

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{3}{4}.$$

Example 7.24.

The probability is 0.6 that Mr. A will pass the examination and the probability is 0.8 that Mr. B will pass the examination. Both sit in the examination. Find the probability that somebody will pass the examination.

Solution:

Both A and B can pass the examination. Therefore, the events A and B are not mutually exclusive.

Hence, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.8 - (0.6)(0.8)$

$= 0.6 + 0.8 - 0.48 = 0.92$

Note: We have assumed that A and B are independent events i.e.,

$$P(A \cap B) = P(A)\,P(B)$$

Example 7.25.

A digit is selected at random from the first 20 digits. Let A be the event that the digit is multiple of 4 and B be the event that the digit is multiple of 6. Find the probability (i) $A \cap B$ (ii) $A \cup B$.

Solution:

Here, $S = \{1, 2, 3, ..., 20\}$, $n(S) = 20$

$A = \{4, 8, 12, 16, 20\}$, $n(A) = 5$

$B = \{6, 12, 18\}$, $n(B) = 3$

$A \cap B = \{12\}$, $n(A \cap B) = 1$

$$P(A) = \frac{n(A)}{n(S)} = \frac{5}{20} \qquad P(B) = \frac{n(B)}{n(S)} = \frac{3}{20}$$

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{20}$$

Thus,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{5}{20} + \frac{3}{20} - \frac{1}{20} = \frac{7}{20}$$

Alternative Method

$S = \{1, 2, 3, ..., 20\}$

$A = \{4, 8, 12, 16, 20\}$, $B = \{6, 12, 18\}$,

$A \cup B = \{4, 6, 8, 12, 16, 18, 20\}$

According to classical definition of probability

$$P(A \cup B) = \frac{\text{Number of points in } A \cup B}{\text{Number of points in } S} = \frac{7}{20}$$

Example 7.26

The following contingency table is set up:

	B	$\bar{B}$
A	10	30
$\bar{A}$	25	35

Determine the probability of :

(i) event A (ii) event B (iii) event $\bar{A}$

(iv) event A and B (v) event A and $\bar{B}$ (vi) event $\bar{A}$ and $\bar{B}$

(vii) event A or B (viii) event A or $\bar{B}$ (ix) event $\bar{A}$ or $\bar{B}$

(x) event $\bar{A}$ or B

Solution:

	B	$\bar{B}$	Total
A	$n(A \cap B) = 10$	$n(A \cap \bar{B}) = 30$	$n(A) = 40$
$\bar{A}$	$n(\bar{A} \cap B) = 25$	$n(\bar{A} \cap \bar{B}) = 35$	$n(\bar{A}) = 60$
Total	$n(B) = 35$	$n(\bar{B}) = 65$	$n(S) = 100$

(i) $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{40}{100} = 0.40$

(ii) $P(B) = \dfrac{n(B)}{n(S)} = \dfrac{35}{100} = 0.35$

(iii) $P(\bar{A}) = \dfrac{n(\bar{A})}{n(S)} = \dfrac{60}{100} = 0.60$

(iv) $P(A \text{ and } B) = \dfrac{n(A \cap B)}{n(S)} = \dfrac{10}{100} = 0.10$

(v) $P(A \text{ and } \bar{B}) = \dfrac{n(A \cap \bar{B})}{n(S)} = \dfrac{30}{100} = 0.30$

(vi) $P(\bar{A} \text{ and } \bar{B}) = \dfrac{n(\bar{A} \cap \bar{B})}{n(S)} = \dfrac{35}{100} = 0.35$

(vii) $P(A \text{ or } B) = P(A) + P(B) - P(A \cap B) = \dfrac{40}{100} + \dfrac{35}{100} - \dfrac{10}{100} = \dfrac{65}{100} = 0.65$

(viii) $P(A \text{ or } \bar{B}) = P(A) + P(\bar{B}) - P(A \cap \bar{B}) = \dfrac{40}{100} + \dfrac{65}{100} - \dfrac{30}{100} = \dfrac{75}{100} = 0.75$

(ix) $P(\bar{A} \text{ or } \bar{B}) = P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cap \bar{B}) = \dfrac{60}{100} + \dfrac{65}{100} - \dfrac{35}{100} = \dfrac{90}{100} = 0.90$

(x) $P(\bar{A} \text{ or } B) = P(\bar{A}) + P(B) - P(\bar{A} \cap B) = \dfrac{60}{100} + \dfrac{35}{100} - \dfrac{25}{100} = \dfrac{70}{100} = 0.70$

Example 7.27

A and B can solve 70 % and 80 % of the problems in a book respectively. Find the chance that a problem chosen at random will be solved by at least one of them.

Solution:

$P(A) = \dfrac{70}{100} = 0.7$, $P(B) = \dfrac{80}{100} = 0.8$ and $P(A \cap B) = P(A)\,P(B) = 0.7 \times 0.8 = 0.56$

($\because$ A and B are independent)

Since A and B are not mutually exclusive events, therefore

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7 + 0.8 - 0.56 = 0.94$

Example 7.28

A class contains 10 boys and 20 girls. Half of the boys and half of the girls have brown eyes. Find the probability that a student chosen at random is a boy or has brown eyes.

Solution:

	Boys	Girls	Total
Class	10	20	30
Brown eyes	5	10	15

$n(S) = {}^{30}C_1 = 30$

Let A = a students is a boy and B = a student has brown eyes

$n(A) = {}^{10}C_1 = 10 \qquad P(A) = \frac{n(A)}{n(S)} = \frac{10}{30} = \frac{1}{3}$

$n(B) = {}^{15}C_1 = 15 \qquad P(B) = \frac{n(B)}{n(S)} = \frac{15}{30} = \frac{1}{2}$

$n(A \cap B) = {}^{5}C_1 = 5 \qquad P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{5}{30} = \frac{1}{6}$

Since A and B are not mutually exclusive events, therefore

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2+3-1}{6} = \frac{4}{6} = \frac{2}{3}$$

Example 7.29

A drum contains 50 bolts and 150 nuts. Half of the bolts and half of the nuts are rusted. (i) If one item is chosen at random, what is the probability that it is rusted or is a bolt? (ii) If two items are chosen at random, what is the probability that both are rusted or both are nuts?

Solution:

Total items = 50 bolts + 150 nuts = 200

Rusted items = 25 bolts + 75 nuts = 100

(i) $n(S) = {}^{200}C_1 = 200$

Let A be the event that an item is rusted
and B be the event that an item is a bolt

$n(A) = {}^{100}C_1 = 100 \qquad P(A) = \frac{n(A)}{n(S)} = \frac{100}{200} = \frac{1}{2}$

$n(B) = {}^{50}C_1 = 50 \qquad P(B) = \frac{n(B)}{n(S)} = \frac{50}{200} = \frac{1}{4}$

$n(A \cap B) = {}^{25}C_1 = 25 \qquad P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{25}{200} = \frac{1}{8}$

Since A and B are not mutually exclusive events, therefore

$$P(A \text{ or } B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{5}{8} = 0.625$$

(ii) $n(S) = {}^{200}C_2 = 19900$

Let C be the event that both items are rusted
and D be the event that both items are nuts

$n(C) = {}^{100}C_2 = 4950 \qquad P(C) = \frac{n(C)}{n(S)} = \frac{4950}{19900} = 0.25$

$n(D) = {}^{150}C_2 = 11175 \qquad P(D) = \frac{n(D)}{n(S)} = \frac{11175}{19900} = 0.56$

$n(C \cap D) = {}^{75}C_2 = 2775 \qquad P(C \cap D) = \frac{n(C \cap D)}{n(S)} = \frac{2775}{19900} = 0.14$

Since C and D are not mutually exclusive events, therefore

$$P(C \text{ or } D) = P(C) + P(D) - P(C \cap D) = 0.25 + 0.56 - 0.14 = 0.67$$

7.7. CONDITIONAL PROBABILITY

The probability of an event A is called conditional if it depends upon the occurrence of some other event B. The conditional probability of event A when event B has already occurred is denoted by $P(A/B)$. It can be shown that

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad [P(B) \neq 0]$$

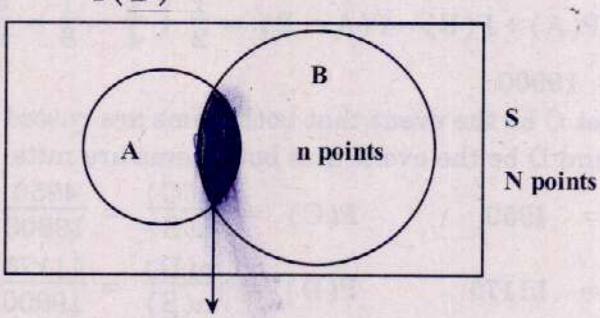
We read $P(A/B)$ as probability of A when B has occurred or "the probability of A given B".

Explanation

The event B has already occurred is called "the additional information" about the sample space S. The additional information reduces the size of the sample space S and the remaining part of S is called the reduced sample space which may be denoted by S_r. Any probability calculated from the reduced sample space is called conditional probability. When the event B has occurred, then the reduced sample space consists of B. Thus $S_r = B$. When it is known that B has occurred, then we are concerned with reduced sample space and not with the original sample space. Suppose a die is rolled and the additional information is that the face is even. Now we are interested to find the probability that the face is a complete square. When it is known that the face is even then the sample space $S = \{1, 2, 3, 4, 5, 6\}$ is reduced to $\{2, 4, 6\}$. Now we are concerned with this reduced sample space. In the original sample space each face has probability of 1/6. Thus in the reduced sample space each face has probability of 1/3. Face 4 is a complete square in the reduced sample space. Thus probability of a complete square when it is known that face is even is 1/3. It is called conditional probability of a complete square when the face is even. If B denotes the even face and A denotes the "complete square" then $P(A/B) = 1/3$.

We can show that the conditional probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) \neq 0$$



$A \cap B$ has m points

Fig. 7.6.

Suppose there is a sample space S which contains N outcomes which are equally likely. The event B has n points and there are m points which are common to the event A and B. Thus $A \cap B$ contains m points. When it is known that B has

occurred, then occurrence of A is possible only when any point out of m points occurs. By using classical definition of probability, we have

$P(A/B) = \frac{m}{n}$ (As B has occurred, therefore B becomes sample space for A)

When both numerator and denominator are divided by N, we get

$$P(A/B) = \frac{m}{n} = \frac{m/N}{n/N} = \frac{P(A\cap B)}{P(B)} \qquad [P(B)\neq 0]$$

when $P(B) = 0$, then $P(A/B)$ does not exist.

When $A\cap B = \phi$, $P(A\cap B) = 0$, therefore $P(A/B) = 0$

Similarly, $P(B/A) = \frac{P(A\cap B)}{P(A)} \qquad [P(A)\neq 0]$

When $A\cap B = \phi$, $P(A\cap B) = 0$ and $P(B/A) = 0$

Thus any conditional probability will be greater than zero only when $A\cap B \neq \phi$, which means that A and B are not mutually exclusive events.

Example 7.30.

A fair die is rolled. Find the probability that the face is even given that the face is less than 4.

Solution:

We have to calculate the conditional probability that the face is even when it is known that the face is less than 4. We can find this conditional probability by two different methods.

Method I: By Sample Space Approach

When a die is tossed, the sample space of 6 equally likely outcomes is,

$$S = \{1, 2, 3, 4, 5, 6\}$$

Let A denote the event "face is even" and B denote the event "face is less than 4"

Then $A = \{2, 4, 6\}$ $B = \{1, 2, 3\}$ and $A\cap B = \{2\}$

From the classical definition of probability, we have,

$$P(B) = \frac{n(B)}{n(S)} = \frac{3}{6} \qquad P(A\cap B) = \frac{n(A\cap B)}{n(S)} = \frac{1}{6}$$

$$P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Which is the required conditional probability.

Method II: By Reduced Sample Space

When a die is tossed, the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

When it is given that the face is less than 4, then the reduced sample space is

$$S_r = \{1, 2, 3\}, \quad n(S_r) = 3$$

Let A = Face is even, $n(A) = 1$, $P(A) = \frac{n(A)}{n(S_r)} = \frac{1}{3}$

Example 7.31.

The workers in a factory are divided into 4 groups as shown below. A worker is selected at random, find the probability that the worker is skilled when it is given that the worker is a male.

	Males	Females
Skilled	80	30
Unskilled	40	10

Solution:

Let A denote the event "worker is skilled" and B denote the event "worker is a male", then A contains 110 workers and B contains 120 workers.

	B	$\bar{B}$	
A	80 $A \cap B$	30	
$\bar{A}$	40	10	
	120		160

There are 80 workers who are males and skilled. Thus $A \cap B$ contains 80 workers.

$$\text{Hence,} \quad P(B) = \frac{\text{Number of workers in B}}{\text{Total number of workers}} = \frac{120}{160} \quad \text{and} \quad P(A \cap B) = \frac{80}{160}$$

$$\text{Hence,} \quad P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{80/160}{120/160} = \frac{80}{120} = \frac{2}{3}$$

Alternate Solution

The conditional probability $P(A / B)$ can be very easily calculated from the idea of reduced sample space. When it is given that the worker is male then the reduced sample space is

	Males
Skilled	80
Unskilled	40
Total	120

In the reduced sample space, 80 workers are skilled. Thus the required conditional probability is 80/120 = 2/3 which is same as calculated earlier.

Example 7.32.

Two dice are rolled. Find the probability that the total on the two dice is less than 5 given that both the faces of dice are similar.

Solution:

When two dice are rolled, there are 36 possible outcomes as given in Table 7.1. All the 36 outcomes are assumed to be equally likely, each having probability of 1/36.

Let A be the event "total on the two dice is less than 5" and B be the event "both faces are similar".

Then $A = \{(1, 1), (1, 2), (2, 1), (1, 3), (3, 1), (2, 2)\}$, $n(A) = 6$

$B = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$, $n(B) = 6$

and $A \cap B = \{(1, 1), (2, 2)\}$, $n(A \cap B) = 2$

Thus $P(B) = \dfrac{n(B)}{n(S)} = \dfrac{6}{36} = \dfrac{1}{6}$

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{2}{36} = \frac{1}{18}$$

Hence, $P(A / B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/18}{1/6} = \dfrac{1}{3}$

Example 7.33.

Given $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$

If $A = \{1, 2, 7\}$ and $B = \{2, 3, 5, 6\}$

Find (i) $P(A / B)$ (ii) $P(B / A)$ (iii) $P(B / \bar{A})$ (iv) $P(\bar{A} / B)$

What relation do you find between $P(A)$ and $P(A / B)$ and $P(B)$ and $P(B / A)$.

Solution:

Given $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$

$A = \{1, 2, 7\}$, $B = \{2, 3, 5, 6\}$

Then $\bar{A} = S - A = \{3, 4, 5, 6, 8\}$, $A \cap B = \{2\}$

$\bar{A} \cap B = \{3, 5, 6\}$

Hence, $P(A) = \dfrac{3}{8}$, $P(B) = \dfrac{4}{8}$, $P(\bar{A}) = \dfrac{5}{8}$

$P(A \cap B) = \dfrac{1}{8}$, $P(\bar{A} \cap B) = \dfrac{3}{8}$

(i) $P(A / B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/8}{4/8} = \dfrac{1}{4}$

Hence, $P(A / B) < P(A)$

(ii) $P(B / A) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{P(A \cap B)}{P(A)} = \dfrac{1/8}{3/8} = \dfrac{1}{3}$

Hence, $P(B / A) < P(B)$

(iii) $P(B / \bar{A}) = \dfrac{P(\bar{A} \cap B)}{P(\bar{A})} = \dfrac{3/8}{5/8} = \dfrac{3}{5}$

Hence, $P(B / \bar{A}) > P(B)$

(iv) $P(\bar{A}/B) = \dfrac{P(\bar{A} \cap B)}{P(B)} = \dfrac{3/8}{4/8} = \dfrac{3}{4}$

Hence, $P(\bar{A}/B) > P(\bar{A})$

In (i) and (ii) above, we find that both the conditional probabilities $P(A/B)$ and $P(B/A)$ are less than the ordinary probabilities and in (iii) and (iv) above, both conditional probabilities are greater than the respective ordinary probabilities.

Thus if A and B are two events then both conditional probabilities $P(A/B)$ and $P(B/A)$ will be greater or both will be less than their ordinary probabilities or both will be equal to their respective ordinary probabilities

If $P(A/B) = P(A)$, then $P(B/A) = P(B)$

The proof of this fact is beyond the scope of this book.

In a topic on Independence we shall say that the events A and B are independent if

$$P(A/B) = P(A) \quad \text{and} \quad P(B/A) = P(B)$$

If $P(A/B) = P(A)$, then $P(B/A)$ is always equal to $P(B)$.

Example 7.34.

Given $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ $A = \{1, 3, 8, 9, 10\}$ $B = \{1, 3, 6, 7\}$ write the elements of the events $A \cap B$ and find the probabilities

(i) $P(A)$ and $P(A/B)$ (ii) $P(B)$ and $P(B/A)$

Solution:

Given $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ $A = \{1, 3, 8, 9, 10\}$ $B = \{1, 3, 6, 7\}$

We have

$$A \cap B = \{1, 3\} \quad \text{and} \quad P(A \cap B) = \frac{2}{10}$$

(i) $P(A) = \dfrac{5}{10} = \dfrac{1}{2}$, $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{2/10}{4/10} = \dfrac{2}{4} = \dfrac{1}{2}$

Thus $P(A) = P(A/B)$

(ii) $P(B) = \dfrac{4}{10} = \dfrac{2}{5}$, $P(B/A) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{2/10}{5/10} = \dfrac{2}{5}$

Hence, $P(B) = P(B/A)$. Here, $P(A/B) = P(A)$ and $P(B/A) = P(B)$.

This is a condition of independence which will be discussed later.

Example 7.35.

Two coins are tossed. Find the following probabilities.

(i) both are heads given that at least one is head.

(ii) the first coin shows head when it is known that one is head and the other is tail.

Solution:

When two coins are tossed the sample space of 4 possible outcomes is

$$S = \{(HH), (HT), (TH), (TT)\}, \; n(S) = 4$$

It is assumed that the four outcomes are equally likely and each outcome has probability of 1/4.

(i) Let A denote the event "both are heads" and B denote the event "at least one is head" then

$$A = \{(HH)\},\ n(A) = 1$$
$$B = \{(HH), (HT), (TH)\},\ n(B) = 3$$
$$A \cap B = \{(HH)\},\ n(A \cap B) = 1$$
$$P(B) = \frac{n(B)}{n(S)} = \frac{3}{4} \text{ and } P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{4}$$

Hence, $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/4}{3/4} = \dfrac{1}{3}$

Alternately

When it is given that at least one is head, then the reduced sample space is $S_r = \{(HH), (HT), (TH)\}$, $n(S_r) = 3$

Let A = Two heads, $n(A) = 1$, $P(A) = \dfrac{n(A)}{n(S_r)} = \dfrac{1}{3}$

(ii) Let A denote the event "the first coin shows head" and B denote the event "one is head and the other is tail" then

$$A = \{(HH), (HT)\},\ n(A) = 2$$
$$B = \{(HT), (TH)\},\ n(B) = 2$$
$$A \cap B = \{(HT)\},\ n(A \cap B) = 1$$

Thus, $P(B) = \dfrac{n(B)}{n(S)} = \dfrac{2}{4}$ and $P(A \cap B) = \dfrac{n(A \cap B)}{n(S)} = \dfrac{1}{4}$

Hence, $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/4}{2/4} = \dfrac{1}{2}$

Alternately

When it is known that one is head and the other is tail, then the sample space S is reduced to $\{(HT), (TH)\}$, $n(S_r) = 2$

Let A = First coin shows head, $n(A) = 1$, $P(A) = \dfrac{n(A)}{n(S_r)} = \dfrac{1}{2}$

In the reduced sample space one outcome is favourable to the event A. Thus the required conditional probability is 1/2.

Example 7.36.

Three babies are born in a family. Find the following probabilities:

(i) a family has 3 male babies given that all the 3 babies have the same gender.

(ii) a family has 2 male babies when it is known that the first baby is male.

(iii) a family has 2 girls given that the middle baby is a male.

(iv) a family has 2 male babies given that the first and last babies have the same gender.

Solution:

It is assumed here that probability of male birth is equal to that of female birth. There are $2^3 = 8$ possible outcomes in the sample space S. Each outcome has probability of 1/8. A boy is denoted by B and a girl is denoted by G. The sample space is

$$S = \{(BBB), (GBB), (BGB), (GGB), (BBG), (GBG), (BGG), (GGG)\}$$

$$n(S) = 8$$

(i) Let A denote the event "3 male babies" and B denote the event "3 babies have the same gender". then

$$A = \{(BBB)\}, \quad n(A) = 1$$

$$B = \{(BBB), (GGG)\}, \quad n(B) = 2$$

$$A \cap B = \{(BBB)\}, \quad n(A \cap B) = 1$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{2}{8} \quad \text{and} \quad P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{8}$$

$$\text{Hence, } P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/8}{2/8} = \frac{1}{2}$$

which is the required conditional probability.

(ii) Let A denote the event "2 male babies" and B denote the event "first baby is male" then

$$A = \{(GBB), (BGB), (BBG)\}, \quad n(A) = 3$$

$$B = \{(BBB), (BGB), (BBG), (BGG)\}, \quad n(B) = 4$$

$$A \cap B = \{(BGB), (BBG)\}, \quad n(A \cap B) = 2$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{8} \quad \text{and} \quad P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{2}{8}$$

$$\text{Hence, } P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{1}{2}$$

(iii) Let A denote the event "2 girls" and B denote the event "middle baby is a male" then $A = \{(GGB), (GBG), (BGG)\}, \; n(A) = 3$

$$B = \{(BBB), (GBB), (BBG), (GBG)\}, \quad n(B) = 4$$

$$A \cap B = \{(GBG)\}, \quad n(A \cap B) = 1$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{8} \quad \text{and} \quad P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{8}$$

$$\text{Thus } P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/8}{4/8} = \frac{1}{4}$$

(iv) Let A denote the event "2 male babies" and B denote the event "first and last babies have the same gender" then

$$A = \{(GBB), (BGB), (BBG)\}, \quad n(A) = 3$$

$$B = \{(BBB), (BGB), (GBG), (GGG)\}, \quad n(B) = 4$$

$$A \cap B = \{(BGB)\}, \quad n(A \cap B) = 1$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{8} \quad \text{and} \quad P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{8}$$

Therefore, $P(A / B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/8}{4/8} = \dfrac{1}{4}$

7.8. INDEPENDENCE OF EVENTS

The two events A and B are called statistically independent if and only if

$$P(A \cap B) = P(A)\,P(B) \qquad P(A) \neq 0,\ P(B) \neq 0$$

A and B are also independent if

$$P(A / B) = P(A) \qquad P(B) \neq 0$$

or $\quad P(B / A) = P(B) \qquad P(A) \neq 0$

In daily life the two events A and B are called independent if they have "nothing to do with each other" or they have no concern with each other. For example one fan is selected from factory A and another fan is taken from factory B. We are interested to find the probability that both are good. To find this probability we use the approach "nothing to do with each other". But this is not the statistical definition of independence. A and B are independent events if and only if

$$P(A \cap B) = P(A)\,P(B).$$

7.9. MULTIPLICATION LAW FOR INDEPENDENT EVENTS

Theorem:

If A and B are independent events having nonzero probabilities, then

$$P(A \cap B) = P(A)\,P(B)$$

The probability $P(A \cap B)$ is called joint probability of A and B and $P(A)$ and $P(B)$ are called the marginal probabilities of A and B respectively.

Proof:

Suppose a sample space S_1 has n total points. The event A belongs to S_1 and has m favourable points. Similarly the sample space S_2 has N total points. The event B belongs to S_2 and has M favourable points.

Therefore, $P(A) = \dfrac{m}{n}$, $\quad P(B) = \dfrac{M}{N}$

Since A and B are independent events, the total points for the combined event A and B will be nN and total favourable points for joint event $A \cap B$ will be mM. Hence

$$P(A \cap B) = \frac{mM}{nN} = \frac{m}{n} \times \frac{M}{N} = P(A)\,P(B)$$

This is called multiplication theorem of probability for independent events. It can be generalized for more than two independent events. For k independent events,

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_k) = P(A_1)\,P(A_2)\,P(A_3) \ldots P(A_k)$$

Corollary: If A and B are independent, then the following equations are also true.

(i) $P(A \cap \bar{B}) = P(A)\,P(\bar{B})$

(ii) $P(\bar{A} \cap B) = P(\bar{A})\,P(B)$

(iii) $P(\bar{A} \cap \bar{B}) = P(\bar{A})\,P(\bar{B})$

Example 7.37.

Two coins are rolled. Show that the event "head on the first coin" and the event "both faces are same" are independent.

Solution:

When two coins are tossed, the sample space is

$$S = \{HH,\ HT,\ TH,\ TT\}$$

The first letter refers to the first coin and the second letter refers to the second coin. Let A denote the event "head on the first coin" and B denote the event "both faces are same". It is assumed that the four outcomes in S are equally likely. Thus probability of 1/4 is assigned to each pair. We have

$$A = \{HH,\ HT\},\quad P(A) = \frac{2}{4} = \frac{1}{2}$$

$$B = \{HH,\ TT\},\quad P(B) = \frac{2}{4} = \frac{1}{2}$$

$$A \cap B = \{HH\},\quad P(A \cap B) = \frac{1}{4}$$

$$P(A \cap B) = P(A)\,P(B) = 1/4$$

Thus A and B are independent.

Note: The students are advised to verify the following equations.

(i) $P(A \cap \bar{B}) = P(A)\,P(\bar{B})$ (ii) $P(\bar{A} \cap B) = P(\bar{A})\,P(B)$

(iii) $P(\bar{A} \cap \bar{B}) = P(\bar{A})\,P(\bar{B})$

Example 7.38.

A six faced die is rolled. Show that the event "face is even" and the event "face is more than 4" are independent.

Solution:

When a die is rolled, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$

Let A denote the event "face is even" and B denote the event "face is more than 4" we have

$$A = \{2, 4, 6\},\ P(A) = \frac{3}{6} = \frac{1}{2}$$

$$B = \{5, 6\},\quad P(B) = \frac{2}{6} = \frac{1}{3}$$

$$A \cap B = \{,6\}, \quad P(A \cap B) = \frac{1}{6}$$

$$P(A \cap B) = P(A)\,P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

Hence A and B are independent events.

Example 7.39.

The probability is 2/3 that Mr. A will pass the examination and the probability is 3/4 that Mr. B will pass the examination. Find the following probabilities

(i) both will pass (ii) only one will pass

(iii) somebody will pass (at least one will pass).

Solution:

Let the symbols A and B denote "passing the examination" and $\bar{A}$ and $\bar{B}$ denote "failing the examination".

We are given $P(A) = \frac{2}{3}$ and $P(B) = \frac{3}{4}$

We have $P(\bar{A}) = 1 - \frac{2}{3} = \frac{1}{3}$ and $P(\bar{B}) = 1 - \frac{3}{4} = \frac{1}{4}$

We shall assume in this case that Mr. A may or may not pass, but the result of Mr. B will not be influenced by the result of Mr. A. Thus A and B are independent. Similarly A and $\bar{B}$, $\bar{A}$ and B, $\bar{A}$ and $\bar{B}$ are also independent.

(i) $P(\text{both will pass}) = P(A \cap B) = P(A) \,.\, P(B) = \frac{2}{3} \times \frac{3}{4} = \frac{1}{2}$

(ii) Only one will pass. This event includes "A pass and B fail" and "A fail and B pass".

$$\begin{aligned}
\text{Thus } P(\text{only one will pass}) &= P(A \cap \bar{B}) + P(\bar{A} \cap B) \\
&= P(A)\,.\,P(\bar{B}) + P(\bar{A})\,.\,P(B) \\
&= \frac{2}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{3}{4} = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}
\end{aligned}$$

$$\begin{aligned}
\text{(iii)} \quad P(\text{somebody will pass}) &= P(\text{at least one will pass}) \\
&= P(A \cap \bar{B}) + P(\bar{A} \cap B) + P(A \cap B) \\
&= P(A)\,P(\bar{B}) + P(\bar{A})\,P(B) + P(A)\,P(B) \\
&= \left(\frac{2}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{4}\right) + \left(\frac{2}{3}\right)\left(\frac{3}{4}\right) = \frac{11}{12}
\end{aligned}$$

Another Approach

The complement of "somebody will pass" is that "both will fail". We know that $P(\text{somebody will pass}) + P(\text{no body will pass}) = 1$

Hence P(somebody will pass) = $1 - P(\text{ no body will pass }) = 1 - P(\bar{A} \cap \bar{B})$

$$= 1 - P(\bar{A})\,P(\bar{B}) = 1 - \frac{1}{3} \times \frac{1}{4} = \frac{11}{12}$$

This result is the same as obtained above.

Example 7.40.

Two cards are drawn from a well-shuffled ordinary pack of 52 cards. The first card is replaced before selecting the second card. Find the following probabilities.

(i) both are kings
(ii) first is king and second is queen
(iii) one is king and one is queen
(iv) both are faced cards
(v) first is card of diamonds and the second is an ace.

Solution:

We shall assume here that all the 52 cards are equally likely and each card has the probability of 1/52. As the first card is returned to the main pack before selecting the second card, therefore the selection of the second card is independent of the first card. Thus the two cards are selected independently.

(i) Both are kings

There are four kings in a playing card, therefore

$$P(\text{first card is a king}) = \frac{4}{52} = P(K_1)$$

$$P(\text{second card is a king}) = \frac{4}{52} = P(K_2)$$

K_1 and K_2 are independent,

$$\text{Thus } P(\text{ both are kings }) = P(K_1 \cap K_2) = P(K_1)\,P(K_2) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}$$

(ii) First is king and second is queen

$$P(\text{ first is king }) = \frac{4}{52} = P(K_1)$$

$$P(\text{ second is queen }) = \frac{4}{52} = P(Q_2)$$

$$P(K_1 \cap Q_2) = P(K_1)\,P(Q_2) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}$$

(iii) One is king and one is queen

This event is different from the event "first is king and the second is queen".

We have $P(\text{ first is king }) = \frac{4}{52} = P(K_1)$

$$P(\text{ second is queen }) = \frac{4}{52} = P(Q_2)$$

$$P(K_1 \cap Q_2) = P(K_1)\,P(Q_2) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}$$

Similarly

$$P(\text{ first is queen }) = \frac{4}{52} = P(Q_1)$$

$$P(\text{ second is king }) = \frac{4}{52} = P(K_2)$$

and $\quad P(Q_1 \cap K_2) = P(Q_1)\,P(K_2) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}$

"One is king and one is queen" includes the two events $(K_1 \cap Q_2)$ and $(Q_1 \cap K_2)$.

Thus $P(\text{ one king, one queen }) = P(K_1 \cap Q_2) + P(Q_1 \cap K_2) = \frac{1}{169} + \frac{1}{169} = \frac{2}{169}$

(iv) Both are faced cards

There are 12 faced cards in a playing card. Therefore,

$$P(\text{ first card is faced }) = \frac{12}{52} = \frac{3}{13} = P(F_1)$$

$$P(\text{ second card is faced }) = \frac{12}{52} = \frac{3}{13} = P(F_2)$$

$$P(\text{ both are faced cards }) = P(F_1 \cap F_2) = P(F_1)\,P(F_2) = \frac{3}{13} \times \frac{3}{13} = \frac{9}{169}$$

(v) First is card of diamonds and the second is an ace

There are 13 cards of diamonds and 4 are aces

$$P(\text{ first card is diamonds }) = \frac{13}{52} = \frac{1}{4} = P(D_1)$$

$$P(\text{ second card is an ace }) = \frac{4}{52} = \frac{1}{13} = P(A_2)$$

$$P(D_1 \cap A_2) = P(D_1)\,P(A_2) = \frac{1}{4} \times \frac{1}{13} = \frac{1}{52}$$

Example 7.41

Three groups of children contain respectively 3 girls and 1 boy; 2 girls and 2 boys; 1 girl and 3 boys. One child is selected at random from each group. Find the chance that the three selected consist of 1 girl and 2 boys.

Solution:

	Girls	Boys	Total
Group 1	3	1	4
Group 2	2	2	4
Group 3	1	3	4

Let G_1 = a girl is selected from group 1 $\quad P(G_1) = 3/4$

G_2 = a girl is selected from group 2 $\quad P(G_2) = 2/4$

G_3 = a girl is selected from group 3 $\quad P(G_3) = 1/4$

B_1 = a boy is selected from group 1 $\quad P(B_1) = 1/4$

B_2 = a boy is selected from group 2 $\quad P(B_2) = 2/4$

B_3 = a boy is selected from group 3 $\quad P(B_3) = 3/4$

We have to find $P(G_1 \cap B_2 \cap B_3) \cup P(B_1 \cap G_2 \cap B_3) \cup P(B_1 \cap B_2 \cap G_3)$

$$= P(G_1)P(B_2)P(B_3) + P(B_1)P(G_2)P(B_3) + P(B_1)P(B_2)P(G_3)$$

$$= \left(\frac{3}{4}\right)\left(\frac{2}{4}\right)\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{2}{4}\right)\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{2}{4}\right)\left(\frac{1}{4}\right)$$

$$= \frac{18}{64} + \frac{6}{64} + \frac{2}{64} = \frac{26}{64} = \frac{13}{32}$$

Example 7.42

For two independent events A and B, $P(A) = \alpha$ and $P(A \cup B) = \beta$ such that $\beta > \alpha$. Then show that $P(B) = \dfrac{\beta - \alpha}{1 - \alpha}$.

Solution:

Since A and B are independent events, therefore

$$P(A \cup B) = P(A) + P(B) - P(A)P(B)$$

$$\beta = \alpha + P(B) - \alpha P(B) \quad \text{or} \quad \beta = \alpha + P(B)(1-\alpha)$$

$$P(B)(1-\alpha) = \beta - \alpha \quad \text{or} \quad P(B) = \frac{\beta - \alpha}{1 - \alpha}$$

7.10. MULTIPLICATION LAW FOR DEPENDENT EVENTS

Theorem:

If A and B are two dependent events, then

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$$

Proof:

Suppose there is a sample space S containing n points (outcomes) which are equally likely. Let event A has m_1 favourable points, event B has m_2 favourable points and $A \cap B$ has m_3 favourable points.

Thus, $P(A) = \dfrac{m_1}{n}$, $m_1 > 0$ $\quad P(B) = \dfrac{m_2}{n}$, $m_2 > 0$ $\quad P(A \cap B) = \dfrac{m_3}{n}$

$$P(A \cap B) = \frac{m_3}{n} = \frac{m_3}{n} \times \frac{m_1}{m_1} = \frac{m_1}{n} \times \frac{m_3}{m_1}$$

$$= \frac{m_1}{n} \times \frac{m_3/n}{m_1/n} = P(A)\,\frac{P(A \cap B)}{P(A)}$$

$$= P(A)P(B/A) \quad \left[\text{Since } P(B/A) = \frac{P(A \cap B)}{P(A)}\right]$$

Similarly,

$$P(A \cap B) = \frac{m_3}{n} = \frac{m_3}{n} \times \frac{m_2}{m_2} = \frac{m_2}{n} \times \frac{m_3}{m_2}$$

$$= \frac{m_2}{n} \times \frac{m_3/n}{m_2/n} = P(B)\,\frac{P(A \cap B)}{P(B)}$$

$$= P(B)P(A/B) \quad \left[\text{Since } P(A/B) = \frac{P(A \cap B)}{P(B)}\right]$$

Hence, $P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$

This is called the general rule of multiplication for probabilities.

Corollary: This rule can be extended to more than two events. For three events named A, B and C, we have ;

$$P(A \cap B \cap C) = P(A)\,P(B/A)\,P(C/A \cap B)$$

If A, B and C are independent, then $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

Example 7.43.

A bag contains two defective D_1, D_2 and three good G_1, G_2, G_3 bulbs. Two bulbs are taken. Find the following probabilities.

(i) First bulb is defective and second is good

(ii) One is defective and one is good.

(iii) Both are good or both defective.

Solution:

(i) Let A represent the event "first is defective" and B represent the event "second is good".

Clearly, P(first defective) = P(A) = 2/5

We assume as if only one bulb has been selected. Now, we assume that event A has occurred. This means that one defective bulb has been taken out from the bag and the bag now has one defective and 3 good bulbs. Another bulb is selected from this bag.

Thus P(second is good/first was defective) = P(B / A) = 3/4

We want the probability that first is defective and second is good. We apply the multiplication rule

$$P(A \cap B) = P(A)\,.\,P(B/A) = \frac{2}{5} \times \frac{3}{4} = \frac{3}{10}$$

(ii) We write the probabilities as if the bulbs have been selected one by one.

P(first defective) = P(A) = 2/5

P(second good/first was defective) = P(B / A) = 3/4

$$P(A \cap B) = P(A)\,P(B/A) = \frac{2}{5} \times \frac{3}{4} = \frac{3}{10} = P(E_1)$$

Now, P(first is good) = 3/5 = P(C)

P(second defective/first was good) = 2/4 = P(D / C)

$$P(C \cap D) = P(C)\,.\,P(D/C) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10} = P(E_2)$$

One defective and one good includes "first defective and second good" and "first good and second defective". From addition law, we get

$$P(\text{1 defective, 1 good}) = P(A \cap B) + P(C \cap D) = P(E_1) + P(E_2)$$

E_1 and E_2 are mutually exclusive events

$$P(E_1 \text{ or } E_2) = P(E_1 \cup E_2) = P(E_1) + P(E_2) = \frac{3}{10} + \frac{3}{10} = \frac{6}{10} = \frac{3}{5}$$

(iii) P(first is good)= 3/5 = $P(A_1)$

P(second good/first was good) $= P(A_2/A_1) = 2/4$

$$P(A_1 \cap A_2) = P(A_1) . P(A_2/A_1) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10} = P(E_3)$$

P(first defective) = $\frac{2}{5} = P(A_3)$

P(second defective/first was defective) $= 1/4 = P(A_4)$

$$P(A_3 \cap A_4) = P(A_3) P(A_4/A_3) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10} = P(E_4)$$

"Both good or both defective" includes E_3 and E_4. E_3 and E_4 are mutually exclusive events in the sample space of two bulbs. We apply addition law of mutually exclusive events.

Thus P(both good or both defective)

$$= P(E_3 \cup E_4) = P(E_3) + P(E_4) = \frac{3}{10} + \frac{1}{10} = \frac{4}{10} = \frac{2}{5}$$

Example 7.44.

In a certain teaching department, there are 4 lecturers and 4 assistant professors. Any three teachers are called upon to give a lecture turn by turn. Find the probability that a lecturer will come first and then the two assistant professors will come.

Solution:

Let A represent the event that first to deliver lecture is a lecturer, B represent the event that the second teacher is assistant professor and let C represent the event that the third teacher is also assistant professor. We have to calculate the probability of the joint event $A \cap B \cap C$ by the multiplication rule

$$P(A \cap B \cap C) = P(A) . P(B/A) P(C/A \cap B)$$

Now $P(A) = 4/8$, $P(B/A) = 4/7$, $P(C/A \cap B) = 3/6$

Hence $P(A \cap B \cap C) = \frac{4}{8} \times \frac{4}{7} \times \frac{3}{6} = \frac{1}{7}$

Example 7.45

What is the probability of getting two consecutive aces in two cards drawn at random from an ordinary deck of 52 playing cards if

(i) the first card is replaced before the second card is drawn

(ii) the first card is not replaced before the second card is drawn

Solution:

(i) $n(S) = {}^{52}C_1 = 52$

Let A be the event that first card drawn is an ace

$n(A) = {}^{4}C_1 = 4$ $\quad P(A) = \frac{n(A)}{n(S)} = \frac{4}{52} = \frac{1}{13}$

Let B be the event that second card drawn is also an ace

$n(B) = {}^4C_1 = 4 \qquad P(B) = \dfrac{n(B)}{n(S)} = \dfrac{4}{52} = \dfrac{1}{13}$

Since A and B are independent events, therefore

$P(A \cap B) = P(A)\,P(B) = \dfrac{1}{13} \times \dfrac{1}{13} = \dfrac{1}{169}$

(ii) $n(S) = {}^{52}C_1 = 52$

Let C be the event that first card drawn is an ace

$n(C) = {}^4C_1 = 4 \qquad P(C) = \dfrac{n(C)}{n(S)} = \dfrac{4}{52} = \dfrac{1}{13}$

Because the first card drawn is not replaced, therefore

$n(S) = {}^{51}C_1 = 51$

Let D be the event that second card drawn is also an ace given that first card was an ace

$n(D/C) = {}^3C_1 = 3 \qquad P(D/C) = \dfrac{n(D/C)}{n(S)} = \dfrac{3}{51} = \dfrac{1}{17}$

Since C and D are dependent events, therefore

$P(C \cap D) = P(C)\,P(D/C) = \dfrac{1}{13} \times \dfrac{1}{17} = \dfrac{1}{221}$

7.11. SELECTIONS WITHOUT AND WITH REPLACEMENT

The students are advised to clearly understand the difference between two types of selections of some elements from a certain lot. The lot may contain any material like bulbs, students, balls, cards etc. There are two methods of selecting a sample from a certain lot. One is called without replacement and the other is called with replacement. When only one element is to be selected from the lot, there is no difference in these two terms. When the second element is selected, the question is whether the first element has been returned to the main lot or not. If the first element is not returned to the main lot before selecting the second element, then the selection is called without replacement. Similarly the second element is not returned before taking the third element and so on. There is another special case in which the first element is returned to the main lot before selecting the next individual. This is called selection with replacement.

7.11.1. WITHOUT REPLACEMENT

There are three methods of calculating the probabilities when a sample is selected from a certain lot without replacement. We explain these methods with the help of a simple example.

Example 7.46.

A container has 3 good (G_1, G_2 and G_3) and 2 defective (D_1 and D_2) items. Two items are selected at random and without replacement. Find the following probabilities.

(i) both are good, (ii) one is good and one is defective.

Solution:

Method I: By Sample Space

This method consists of making the sample space and then writing the probabilities of the required sub-sets. The sample space S consists of all possible combinations. It should be remembered that we shall be required to make the permutations if the event is "first will be good and second will be defective". Here we have 5 objects and 2 objects are selected at a time. Thus the number of possible combinations (outcomes) are

$$^{N}C_{n} = {}^{5}C_{2} = \frac{5!}{2!\,(5-2)!} = 10$$

The sample space of 10 outcomes is

$$S = \{G_1G_2, G_1G_3, G_2G_3, G_1D_1, G_2D_1, G_3D_1, G_1D_2, G_2D_2, G_3D_2, D_1D_2\}$$

(i) Let A denote the event "both are good". Then

$$A = \{G_1G_2, G_1G_3, G_2G_3\}$$

$$P(A) = \frac{\text{Number of points in A}}{\text{Total number of outcomes in S}} = \frac{3}{10}$$

(ii) Let B denote the event "one good and one defective". Then

$$B = \{G_1D_1, G_2D_1, G_3D_1, G_1D_2, G_2D_2, G_3D_2\}$$

$$P(B) = \frac{\text{Number of points in B}}{\text{Total number of outcomes in S}} = \frac{6}{10}$$

It should be remembered that the event "one good and one defective" is also called "exactly one good" or "only one good".

Note: In this question we have made the sample space of combinations. We shall get the same result if we write all possible permutations. This is left for an exercise for the students.

Method II: Using Laws of Probability

In this method, we shall assume that the items are selected one by one. Let A be the event that the item on the first draw is good and B be the event that the item on the second draw is also good.

(i) $P(A) = 3/5, \quad P(B/A) = 2/4$

Using multiplication law of probability, we get

$$P(\text{both good}) = P(A \cap B) = P(A)\,.\,P(B/A) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

Which is the same as obtained in Method I.

(ii) $P(\text{first item is good}) = P(G) = \frac{3}{5}$

$P(\text{second item is defective /first item is good}) = P(D/G) = 2/4$

$$P(\text{first good and second defective}) = P(G \cap D) = P(G)\,P(D/G)$$

$$= \frac{3}{5} \times \frac{2}{4} = \frac{3}{10} = P(E_1)$$

Similarly,

$$P(\text{ first defective and second good }) = P(D \cap G) = P(D).P(G/D) = \frac{2}{5}\times\frac{3}{4} = \frac{3}{10} = P(E_2)$$

We need the probability of the event "one good, one defective". This event includes E_1 and E_2.

$$\text{Therefore, } P(E_1 \text{ or } E_2) = P(E_1 \cup E_2) = P(E_1) + P(E_2) = \frac{3}{10} + \frac{3}{10} = \frac{6}{10}$$

The answer is the same as obtained in Method I.

Method III: By Hypergeometric Distribution

The first method is suitable when the size of the sample space is small and the second method is suitable when the sample size consists of a few elements. Method III can be used when the sample space is very large. We get the required probabilities as

(i) $P(\text{ both good }) = \dfrac{{}^3C_2 \,.\, {}^2C_0}{{}^5C_2} = \dfrac{3}{10}$

(ii) $P(\text{ one good, one defective }) = \dfrac{{}^3C_1 \,.\, {}^2C_1}{{}^5C_2} = \dfrac{6}{10}$

Students are advised to repeat this question when there are 30 good and 20 defective items and two items are selected without replacement

7.11.2. WITH REPLACEMENT

When a sample is selected with replacement, we can calculate the probabilities by three different methods discussed in the example below.

Example 7.47.

A container has 3 good (G_1, G_2 and G_3) items and 2 defective items (D_1 and D_2). Two items are selected with replacement. Find the following probabilities.

(i) both are good,

(ii) one is good and one is defective.

(Example 7.39 has been repeated here for the convenience of students)

Solution:

Method I: By Sample Space

When 2 items are selected out of 5 items with replacement, the sample space S consists of 25 pairs $5^2 = 25$. The sample space is

	G_1	G_2	G_3	D_1	D_2
G_1	G_1G_1	G_1G_2	G_1G_3	G_1D_1	G_1D_2
G_2	G_2G_1	G_2G_2	G_2G_3	G_2D_1	G_2D_2
G_3	G_3G_1	G_3G_2	G_3G_3	G_3D_1	G_3D_2
D_1	D_1G_1	D_1G_2	D_1G_3	D_1D_1	D_1D_2
D_2	D_2G_1	D_2G_2	D_2G_3	D_2D_1	D_2D_2

$= S$

(i) Let A represent the event "both good" then

$$A = \{G_1G_1, G_1G_2, G_1G_3, \ldots, G_3G_3\}.$$ The event A contains 9 outcomes

Therefore, $P(A) = \dfrac{\text{Number of outcomes in A}}{\text{Total number of outcomes in S}} = \dfrac{9}{25}$

(ii) Let B represent the event "one good, one defective", then B has 12 outcomes.

Therefore, $P(B) = 12/25$

Method II: Using Laws of Probability

Here we shall consider that the items are drawn one by one.

(i) Thus P(first item is good) $= P(A) = 3/5$

and P(second item is good) $= P(B) = 3/5$

As the 1st item is returned to the main lot, therefore the second item is selected from the container having the same original contents. We need the probability that both are defective. It is obtained by using multiplication law of independent events.

Thus P(both good) $= P(A \cap B) = P(A) . P(B) = 3/5 \times 3/5 = 9/25$

(ii) P(first is good) $= P(G) = 3/5$

P(second is defective) $= P(D) = 2/5$

$P(G \cap D) = P(G) . P(D)$

$= 3/5 \times 2/5 = 6/25 = P(E_1)$

Similarly, we have

$P(D \cap G) = P(D) . P(G)$

$= 2/5 \times 3/5 = 6/25 = P(E_2)$

The event "one good, one defective" includes E_1 and E_2. As E_1 and E_2 are mutually exclusive events, therefore

$P(E_1 \text{ or } E_2) = P(E_1 \cup E_2) = P(E_1) + P(E_2) = 6/25 + 6/25 = 12/25$

Method III: By Binomial Distribution

This method is applicable in special cases when the main lot can be divided into two seperate groups. For example the good and defective items in a certain lot, intelligent and non-intelligent students in a class, smokers and non-smokers in a city and so on. This method shall be discussed in detail in chapter binomial probability distribution.

SHORT DEFINITIONS

Factorial

Given the positive integer n, the product of all the whole numbers from n down through 1 is called n factorial and is written as n!

or

Factorial is defined to be the product of all natural numbers less than or equal to a particular number.

Permutation

A permutation is an arrangement in which the order of the objects selected from a specific pool of objects is important.

or

A permutation is an ordered arrangement of objects.

Combination

A combination is collection of a group of objects without regard to order.

or

A combination is an arrangement of objects without regard to order.

Experiment

An experiment is the observation of some activity or the act of taking some measurement.

or

An experiment is any process that allows researchers to obtain observations.

Random Experiment

A random experiment is an experiment where there is uncertainty about the occurrence of an outcome.

or

A random experiment is any well-defined procedure that produces an observable outcome that could not be perfectly predicted in advance.

Sample Point

An individual outcome of an experiment is called a sample point.

Outcome

An outcome is a particular result of an experiment.

or

Each distinct possible result of an experiment is called an outcome.

Sample Space

The collection of all possible outcomes for a random experiment is called the sample space.

or

The sample space for a random experiment is the set of all possible outcomes from that experiment.

Event

An event is the collection of one or more outcomes of an experiment.

or

Any subset of a sample space is called an event.

Simple Event

A single outcome of an experiment is called a simple event.

or

A simple event is an outcome that cannot be broken down any more.

Compound Event

An event combining two or more simple events is called a compound event.

or

If an event contains more than one sample points, it is called a compound event.

Venn Diagram

Diagram used to portray a sample space or a relationship between events is called venn diagram.

or

A venn diagram is a graphical device for representing symbolically the sample space and operations involving events.

Tree Diagram

A method for illustrating the sequence of several events is called a tree diagram.

or

A tree diagram is a graphical device helpful in defining sample points of an experiment involving multiple steps.

Equally Likely Events

Events that have the same probability of occurrence are called equally likely events.

or

Two events are equally likely when there is no reason to expect that one event has more chance to occur than the other event.

Mutually Exclusive Events

Two events are said to be mutually exclusive if they have no outcomes in common.

or

Two or more events that cannot occur together on a single trial of an experiment are mutually exclusive events.

Exhaustive Events

A set of events is said to be collectively exhaustive if their union is equal to sample space S.

Completely Exhaustive Event

An event is said to be completely exhaustive if it is alone equal to the sample space S.

Probability

Probability is the likelihood or chance that a particular event will occur.

or

Probability is a number between zero and one that is used to measure uncertainty.

Basic Properties of Probability

(i) The probability of an event is always between 0 and 1.

(ii) An event which cannot occur has the probability zero.

(iii) The probability of any event that is certain to occur is one.

(iv) The sum of all the associated probabilities must be one.

Classical or Mathematical Definition of Probability

Classical probability is applied when the possible outcomes of an experiment are equally likely to occur. If an experiment can produce N mutually exclusive and equally likely outcomes out of which n outcomes are favourable to the occurrence of event A, then the probability of event A is

$$P(A) = \frac{\text{Number of outcomes favourable to A}}{\text{Total number of possible outcomes}} = \frac{n}{N}$$

Relative Frequency or Statistical or Empirical Definition of Probability

This definition is based on the actual observations when the random experiment is repeated a large number of times under similar conditions. If an experiment is repeated N times where N is large and the event A is observed n times, the probability of A is

$$P(A) = \underset{N \to \infty}{\text{Limit}} \frac{n}{N}$$

Subjective Definition of Probability

According to this definition, the probability of an event is the degree of confidence or belief on the part of the statistician or decision maker that the event will occur.

Conditional Probability

A conditional probability is the probability that an event will occur given that another event has already occurred. ***or***

The probability associated with the reduced sample space is called conditional probability.

Independent Events

Two events are said to be independent if the occurrence or nonoccurrence of one event does not affect the probability that the other will occur. ***or***

Two events A and B are independent if the probability of the second event B is not affected by the occurrence or nonoccurrence of the first event A.

Dependent Events

Two events are dependent if the occurrence of one event impacts the probability of the occurrence of the other event. ***or***

Two events are said to be dependent events if occurrence of one event affects the occurrence of other event.

Statistical Independence

If events A and B are statistically independent, the probability of the occurrence of one event is not affected by the occurrence of the other. That is, $P(A / B) = P(A)$ and $P(B / A) = P(B)$.

Statement of Addition Rule for Mutually Exclusive Events

If A and B are two events, the probability that either event A or event B occurs equals the probability that event A occurs plus the probability that event B occurs. In symbols, $P(A \text{ or } B) = P(A) + P(B)$.

Statement of General Addition Rule

Given two events A and B, the probability that event A or event B or both occur is equal to the probability that event A occurs, plus the probability that event B occurs, minus the probability that both events occur. In symbols,

$$P(A \text{ or } B \text{ or both}) = P(A) + P(B) - P(A \cap B).$$

Statement of Multiplication Rule for Independent Events

If A and B are two events, the probability that both event A and event B occur equals the probability that event A occurs into the probability that event B occurs. In symbols, $P(A \text{ and } B) = P(A)\, P(B)$.

Statement of General Multiplication Rule

The probability of the joint occurrence of event A and event B is equal to the probability of A into the conditional probability of B given A or probability of B into the conditional probability of A given B. In symbols,

$$P(A \cap B) = P(A)\, P(B / A) = P(B)\, P(A / B).$$

MULTIPLE – CHOICE QUESTIONS

1. When the possible outcomes of an experiment are equally likely to occur, then we apply:
   (a) relative probability (b) subjective probability
   (c) conditional probability (d) classical probability
2. A number between 0 and 1 that is used to measure uncertainty is called:
   (a) random variable (b) trial
   (c) simple event (d) probability
3. Probability lies between:
   (a) -1 and $+1$ (b) 0 and 1
   (c) 0 and n (d) 0 and ∞
4. Probability can be expressed as:
   (a) ratio (b) fraction
   (c) percentage (d) all of the above
5. The probability based on the concept of relative frequency is called:
   (a) empirical probability (b) statistical probability
   (c) both (a) and (b) (d) neither (a) nor (b)
6. The probability of an event cannot be:
   (a) equal to zero (b) greater than zero
   (c) equal to one (d) less than zero

7. A measure of the chance that an uncertain event will occur:
   (a) an experiment (b) an event
   (c) a probability (d) a trial
8. A graphical device used to list all possibilities of a sequence of outcomes in a systematic way is called:
   (a) probability histogram (b) venn diagram
   (c) pie diagram (d) tree diagram
9. A random experiment contains:
   (a) at least one outcome (b) at least two outcomes
   (c) at most one outcome (d) at most two outcomes
10. The probability of all possible outcomes of a random experiment is always equal to:
   (a) one (b) zero
   (c) infinity (d) all of the above
11. The outcome of tossing a coin is a:
   (a) mutually exclusive event (b) compound event
   (c) certain event (d) simple event
12. The result of no interest of an experiment is called:
   (a) constant (b) event
   (c) failure (d) success
13. A set of all possible outcomes of an experiment is called:
   (a) combination (b) sample point
   (c) sample space (d) compound event
14. The number of counting rules that are useful in determining the number of outcomes in an experiment are:
   (a) one (b) two
   (c) three (d) four
15. The events having no experimental outcomes in common is called:
   (a) equally likely events (b) exhaustive events
   (c) mutually exclusive events (d) Independent events
16. A set of outcomes formed after some additional information is called:
   (a) sample space (b) reduced sample space
   (c) null set (d) random experiment
17. The probability associated with the reduced sample space is called:
   (a) conditional probability (b) statistical probability
   (c) mathematical probability (d) subjective probability
18. An arrangement of objects without regard to order is called:
   (a) permutation (b) combination
   (c) random experiment (d) sample point

19. The number of permutations of a set of n things, taken r at a time with $n \geq r$ is given by:

(a) $\frac{n!}{(n-r)!}$ (b) $\frac{n!}{(r-n)!}$

(c) $\frac{n!}{r!(n-r)!}$ (d) $\frac{n!}{r!}$

20. If three candidates are selected to attend a course from the ten candidates, the number of ways of selecting the candidates is an example of:

(a) combination (b) permutation

(c) reduced sample space (d) both (a) and (b)

21. When each outcome of a sample space is as likely to occur as any other, the outcomes are called:

(a) exhaustive (b) mutually exclusive

(c) equally likely (d) not mutually exclusive

22. If A is any event in S and $\bar{A}$ its complement, then $P(\bar{A})$ is equal to:

(a) 1 (b) 0

(c) 1 – A (d) 1 – P(A)

23. When certainty is involved in a situation, its probability is equal to:

(a) zero (b) between – 1 and + 1

(c) between 0 and 1 (d) one

24. Which of the following cannot be taken as probability of an event?

(a) 0 (b) 0.5

(c) 1 (d) – 1

25. If an event contains more than one sample points, it is called a:

(a) simple event (b) compound event

(c) impossible event (d) certain event

26. When the occurrence of one event has no effect on the probability of the occurrence of another event, the events are called:

(a) independent (b) dependent

(c) mutually exclusive (d) equally likely

27. A particular result of an experiment is called:

(a) trial (b) simple event

(c) compound event (d) outcome

28. A collection of one or more outcomes of an experiment is called:

(a) event (b) outcome

(c) sample point (d) none of the above

29. A process that leads to the occurrence of one and only one of several possible observations is called:

(a) random experiment (b) random variable

(c) experiment (d) probability distribution

30. Which statement is false?
   (a) The classical definition applies when there are n equally likely outcomes to an experiment
   (b) The empirical definition occurs when number of times an event happens is divided by the number of observations.
   (c) A subjective probability is based on whatever information is available
   (d) The general rule of addition is used when the events are mutually exclusive

31. The term 'sample space' is used for:
   (a) all possible outcomes (b) all possible coins
   (c) probability (d) sample.

32. The term 'event' is used for:
   (a) time (b) a sub-set of the sample space
   (c) probability (d) total number of outcomes.

33. The six faces of the die are called equally likely if the die is:
   (a) small (b) fair
   (c) six-faced (d) round.

34. If we toss a coin and $P(H) = 2P(T)$, then probability of head is equal to:
   (a) 0 (b) 1/2
   (c) 1/3 (d) 2/3

35. A letter is chosen at random from the word "Statistics". The probability of getting a vowel is:
   (a) 1/10 (b) 2/10
   (c) 3/10 (d) 4/10

36. An arrangement in which the order of the objects selected from a specific pool of objects is important called:
   (a) combination (b) permutation
   (c) factorial (d) sample space

37. Two books are to be selected at random without replacement out of four books. The number of possible selections are:
   (a) 4 (b) 2
   (c) 6 (d) 3

38. Three books of different colours are to be arranged in a book-shelf. The possible arrangements are:
   (a) 3 (b) 1
   (c) 6 (d) 2

39. If a sample $S = \{1, 2\}$, the number of all possible sub-sets are:
   (a) 2 (b) 1
   (c) 3 (d) 4

40. When a die and a coin are rolled together, all possible outcomes are:
   (a) 6 (b) 2
   (c) 36 (d) 12

41. When two coins are tossed, the possible outcomes are:
(a) 2 (b) 4
(c) 1 (d) none of them.

42. If three coins are tossed, the possible outcomes are:
(a) 8 (b) 3
(c) 1 (d) none of them.

43. If n coins are tossed, the possible outcomes are:
(a) n (b) 2
(c) 2^n (d) all of them.

44. If two dice are rolled, the possible outcomes are:
(a) 6 (b) 36
(c) 1 (d) difficult to answer.

45. When **n** dice are rolled, the possible outcomes are:
(a) 6^n (b) 6
(c) 1 (d) 18.

46. When one card is selected at random from a pack of 52 playing cards, the possible selections are:
(a) 104 (b) 52
(c) 520 (d) 2704

47. Two cards are selected at random with replacement from a pack of 52 playing cards. The possible outcomes are:
(a) 52×52 (b) 52
(c) 1326 (d) 2

48. A bag contains 4 white and 2 black balls of the same size and weight, and two balls are selected at random without replacement, the possible selections are:
(a) 6 (b) 4
(c) 36 (d) 15

49. Two balls are selected at random with replacement from a bag containing 3 red, 3 black and 2 green balls. The possible outcomes are:
(a) 8 (b) 64
(c) 16 (d) 2

50. Five cards are selected at random from a pack of 52 cards with replacement. The possible combinations are:
(a) 52 (b) $(52)^5$
(c) 52×52 (d) $(5)^{52}$

51. The digits 1, 2, 3, 4, 5 are the roll numbers of 5 students. These roll numbers are written on the paper slips and two paper slips are selected at random without replacement. The possible combinations are:
(a) 5 (b) 2
(c) 25 (d) 10

52. Which is the impossible event when a die is rolled?
(a) 2 or 3 (b) 5 or 6
(c) 1 (d) 0 or 7

53. The probability of drawing any one spade card is:
(a) 1/13 (b) 1/4
(c) 4/13 (d) 1/52

54. A balance die is rolled, the probability of getting an odd number is:
(a) 1/2 (b) 1/4
(c) 1/6 (d) 1/3

55. Two fair dice are rolled. The probability of throwing an odd sum is:
(a) 1 (b) 1/2
(c) 1/6 (d) 1/36

56. Given $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cup B) = 0.9$, then:
(a) A and B are not mutually exclusive events
(b) A and B are equally likely events
(c) A and B are independent events
(d) A and B are mutually exclusive events

57. If $P(B/A) = 0.50$ and $P(A \cap B) = 0.40$, then $P(A)$ will be equal to:
(a) 0.40 (b) 0.50
(c) 0.80 (d) 1

58. Which of the following statements is incorrect:
(a) $A - (B \cup C) = (A - B) \cap (A - C)$ (b) $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$
(c) $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$ (d) $A - (B \cap C) = (A + B) \cup (A - C)$

59. If $P(A/B) = P(A)$ and $P(B/A) = P(B)$, then A and B are:
(a) mutually exclusive (b) dependent
(c) equally likely (d) independent

60. A fair coin is tossed 100 times, the expected number of heads is:
(a) 100 (b) 50
(c) 30 (d) 60

61. When two dice are rolled, the maximum total on the two faces of the dice will be:
(a) 6 (b) 36
(c) 12 (d) 2

62. A random sample of 200 random digits is selected from a random number table. Expected number of zeros in the sample is:
(a) zero (b) 10
(c) 20 (d) 5

63. Six digits are selected at random again and again from a random number table and the even digits are counted each time. In most of the cases, the number of even digits will be:
(a) 2 (b) 3
(c) 4 (d) 6

64. Two events A and B are called mutually exclusive if:
(a) $A \cup B = \varphi$ (b) $A \cap B = \varphi$
(c) $A \cap B = S$ (d) $A \cap B = 1$

65. If A and B are two mutually exclusive events, then:
(a) $P(A \cap B) = 0$ (b) $P(A \cap B) = 1$
(c) $P(A \cup B) = 0$ (d) $P(A \cap B) = S$.

66. When A and B are two non-empty and mutually exclusive events, then:
(a) $P(A \cup B) = P(A) . P(B)$ (b) $P(A \cup B) = P(A) + P(B)$
(c) $P(A \cap B) = P(A) . P(B)$ (d) $P(A \cap B) = P(A) + P(B)$

67. The two events A and B are called not mutually exclusive events if:
(a) $A \cap B = \varphi$ (b) $A \cap B \neq \varphi$
(c) $A \cup B = \varphi$ (d) $A \cap B =$ zero

68. If A and B are disjoint events then the statement which is always true is:
(a) $P(A/B) = 0$ (b) $P(A \cup B) = 0$
(c) $P(A \cap B) = 1$ (d) $P(A) = P(B)$

69. The events A, B and C are called exhaustive events if:
(a) $A \cup B \cup C = S$ (b) $A \cap B \cap C = S$
(c) $A \cup B \cup C = \varphi$ (d) $A \cup B \cup C =$ zero.

70. If A and B are not-mutually exclusive events, then:
(a) $P(A \cup B) + P(A \cap B) = P(A) + P(B)$ (b) $P(A \cup B) = P(A) + P(B)$
(c) $P(A \cup B) = P(A) P(B)$ (d) $P(A \cap B) = P(A) + P(B)$

71. If an event $\bar{A}$ is the complement of the event A, then:
(a) $A \cup \bar{A} = S$ (b) $A \cap \bar{A} = S$
(c) $A \cup \bar{A} = \varphi$ (d) $P(A) = P(\bar{A})$

72. If $A_1, A_2, A_3, \cdots, A_k$ are k mutually exclusive events, then:
(a) $P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_k)$
(b) $P(A_1 \cup A_2 \cup \cdots \cup A_k) > 1$
(c) $P(A_1 \cap A_2 \cap \cdots \cap A_k) = 1$
(d) $P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1 \cup A_2 \cup \cdots \cup A_k)$.

73. If A is an empty set and B is a non-empty set then:
(a) $A \cap B = S$ (b) $A \cap B = B$
(c) $A \cup B = B$ (d) $P(A) = P(B)$

74. If A is an empty set (φ) and S is the sample space then:

(a) $P(A \cup S) = P(S)$ (b) $P(A \cup S) = P(\varphi)$

(c) $P(A \cap S) = 1$ (d) $P(A \cup S) = \text{zero}$

75. If A and B are independent events, then:

(a) $P(A \cup B) = P(A)\ P(B)$ (b) $P(A \cap B) = P(A)\ P(B)$

(c) $P(A \cap B) = P(A) + P(B)$ (d) $P(A) = P(B)$

76. If A and B are two independent events, then:

(a) $P(A/B) = P(A)$ (b) $P(A) = P(B)$

(c) $P(A) < P(B)$ (d) $P(A/B) = P(B/A)$.

77. A and B are two independent events. Which one of these equations is false?

(a) $P(A \cap \bar{B}) = P(A)\ P(\bar{B})$ (b) $P(\bar{A} \cap \bar{B}) = P(\bar{B} \cap \bar{A})$

(c) $P(\bar{A} \cap \bar{B}) = P(\bar{A})\ P(\bar{B})$ (d) $P(A \cup B) = P(A)\ P(B)$

78. The conditional probability of the event A when event B has occurred is denoted by:

(a) $P(A + B)$ (b) $P(A - B)$

(c) $P(A/B)$ (d) $P(\bar{A})$

79. If A and B are any two events, then $P(A/B) + P(\bar{A}/B)$ is equal to:

(a) 0 (b) 0.25

(c) 0.5 (d) 1

80. If A is an arbitrary event, then $P(A/A)$ is equal to:

(a) zero (b) one

(c) infinity (d) less than one

81. If A and B are any two events, then $P(\bar{A}/B)$ is equal to:

(a) $P(A/B)$ (b) $1 - P(A/B)$

(c) $1 + P(A/B)$ (d) $P(\bar{A} \cap B)$

82. If A and B are any two events, then $P(A \cup \bar{B})$ is equal to:

(a) $1 + P(A \cap B)$ (b) $1 - P(A \cup B)$

(c) $1 - P(A \cap B)$ (d) $P(A) + P(B)$

83. If A and B are any two events, then $P(\bar{A} \cap \bar{B})$ is equal to:

(a) $1 - P(A \cup B)$ (b) $1 - P(A \cap B)$

(c) $1 - P(\bar{A} \cap B)$ (d) $1 - P(A \cap \bar{B})$

84. Which of the following statements is correct:

(a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(b) $A \cup (B \cap C) = (A \cap B) \cap (A \cup C)$

(c) $A \cap (B \cap C) = (A \cup B) \cup (B \cap C)$

(d) $A \cap (B \cup C) = A + (B \cap C)$

85. If A and B are two mutually exclusive and exhaustive events and $P(A) = 2P(B)$, then $P(B)$ is equal to:

(a) 1/2 (b) 2/3

(c) 1/3 (d) 1/4

86. Two coins are tossed. Probability of getting head on the first coin is:

(a) 2/4 (b) 1

(c) zero (d) 4

87. A die and a coin are tossed together. Probability of getting head on the coin is:

(a) 6/12 (b) 6

(c) 12 (d) zero

88. A fair die is rolled. Probability of getting even face given that face is less than 5 is given by:

(a) 1/2 (b) 5

(c) 2 (d) 6

89. Two coins are tossed. The probability that both faces will be matching is given by:

(a) 1/4 (b) 1/2

(c) 1 (d) zero.

90. Two coins are tossed. Probability of getting two heads given that there is at least one head is given by:

(a) 1/2 (b) 1/3

(c) 1/4 (d) 2/3

91. A fair die is rolled. Probability of getting more than 4 or less than 3 is given by:

(a) 2/3 (b) 1/3

(c) 1/2 (d) 4/3

92. A fair die is rolled. Probability of getting even face or face more than 4 is:

(a) 1/3 (b) 2/3

(c) 1/2 (d) 5/6

93. Two dice are rolled. Probability of getting similar faces is:

(a) 5/36 (b) 1/6

(c) 1/3 (d) 1/2

94. Two dice are rolled. Probability of getting a total less than 4 or total more than 10 is given by:

(a) 10/36 (b) 4/36

(c) 1/6 (d) 14/36

95. Two dice are rolled. Probability of getting a total of 4 given that both faces are similar is:
   (a) 5/36 (b) 1/36
   (c) 4/36 (d) 1/6

96. If A and B are two not-independent events, then the probability that both A and B will happen together is:
   (a) $P(A \cap B) = P(A)\ P(B/A)$ (b) $P(A \cap B) = P(A)\ P(B)$
   (c) $P(A \cap B) = P(A) + P(B)$ (d) $P(A \cap B) = P(A)$

97. If A and B are two dependent events, then:
   (a) $P(A)\ P(B/A) = P(B)\ P(A/B)$ (b) $P(A/B) = P(B/A)$
   (c) $P(A/B) = P(A)$ (d) $P(A) = P(B)$.

98. Which one is true?
   (a) $P(A \cap \bar{B}) = P(B) - P(A \cup B)$ (b) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$
   (c) $P(A \cap \bar{B}) = P(B) - P(A \cap B)$ (d) $P(A \cap \bar{B}) = P(A) + P(\bar{B})$.

99. Given $P(A \cap B) = 3/5$, then $P(\bar{A} \cup \bar{B})$ is:
   (a) 1/5 (b) 2/5
   (c) 3/5 (d) 1

100. Given $P(\bar{A} \cap \bar{B}) = 3/10$, then $P(A \cup B)$ is:
   (a) 7/10 (b) 1/10
   (c) 3/10 (d) 1.

101. Given $P(A) = 2/3$, $P(B) = 3/8$ and $P(A \cap B) = 1/4$, then A and B are:
   (a) independent (b) dependent
   (c) mutually exclusive (d) equally likely

ANSWERS

1.	(d)	2.	(d)	3.	(b)	4.	(d)	5.	(c)	6.	(d)	7.	(c)	8.	(d)
9.	(b)	10.	(a)	11.	(d)	12.	(c)	13.	(c)	14.	(c)	15.	(c)	16.	(b)
17.	(a)	18.	(b)	19.	(a)	20.	(a)	21.	(c)	22.	(d)	23.	(d)	24.	(d)
25.	(b)	26.	(a)	27.	(d)	28.	(a)	29.	(c)	30.	(d)	31.	(a)	32.	(b)
33.	(b)	34.	(d)	35.	(c)	36.	(b)	37.	(c)	38.	(c)	39.	(d)	40.	(d)
41.	(b)	42.	(a)	43.	(c)	44.	(b)	45.	(a)	46.	(b)	47.	(a)	48.	(d)
49.	(b)	50.	(b)	51.	(d)	52.	(d)	53.	(b)	54.	(a)	55.	(b)	56.	(d)
57.	(c)	58.	(d)	59.	(d)	60.	(b)	61.	(c)	62.	(c)	63.	(b)	64.	(b)
65.	(a)	66.	(b)	67.	(b)	68.	(a)	69.	(a)	70.	(a)	71.	(a)	72.	(a)
73.	(c)	74.	(a)	75.	(b)	76.	(a)	77.	(d)	78.	(c)	79.	(d)	80.	(b)
81.	(b)	82.	(c)	83.	(a)	84.	(a)	85.	(c)	86.	(a)	87.	(a)	88.	(a)
89.	(b)	90.	(b)	91.	(a)	92.	(b)	93.	(b)	94.	(c)	95.	(d)	96.	(a)
97.	(a)	98.	(b)	99.	(b)	100	(a)	101	(a)						

SHORT QUESTIONS

Q.1 Explain the random experiment by giving suitable examples.

Q.2 What is a random experiment and what are its properties?

Q.3 Write short note on sample space.

Q.4 What is meant by an event?

Q.5 Distinguish between simple and compound events.

Q.6 Define Venn diagram.

Q.7 Explain the concept of equally likely events and give one example.

Q.8 Explain the concept of mutually exclusive events.

Q.9 Distinguish between mutually exclusive and not mutually exclusive events.

Q.10 Define collectively exhaustive events.

Q.11 Differentiate between combinations and permutations.

Q.12 When are two events A and B said to be independent?

Q.13 Explain the concept of dependent events.

Q.14 Distinguish between independent and dependent events.

Q.15 Why do we study probability theory?

Q.16 Write down a definition of probability.

Q.17 Differentiate between independent and mutually exclusive events.

Q.18 Define the complement.

Q.19 What is the sample space for the roll of a single six-sided die?

Ans. $S = \{1, 2, 3, 4, 5, 6\}$

Q.20 Write down the basic properties of probability

Q.21 Write down the classical definition of probability.

Q.22 Write down the relative frequency definition of probability.

Q.23 Write down the three definitions of probability.

Q.24 Define addition law of probability.

Q.25 Define conditional probability.

Q.26 Define multiplication law of probability.

Q.27 State the addition law of probabilities for mutually exclusive events.

Q.28 State the addition law of probabilities for not mutually exclusive events.

Q.29 Write the statement for multiplication law of probability for independent events.

Q.30 What is meant by conditional probability? Give examples.

Q.31 State the multiplication law of probabilities for dependent events.

Q.32 State the multiplication rule for two independent events, E_1 and E_2.

Q.33 Define the classical method of probability.

Q.34 Define the relative frequency method of probability.

Q.35 Define the subjective method of probability.

Q.36 A fair coin is tossed twice. What is the probability of obtaining exactly one tail?

Ans. 1/2

Q.37 Three coins are tossed. Count the total number of sample points in S.

Ans. 8

Q.38 If three coins are tossed, what is the probability of getting exactly two heads?

Ans. 3/8

Q.39 If three coins are tossed, what is the probability of getting at least two heads?

Ans. 1/2

Q.40 If three coins are tossed, what is the probability of getting at most two heads?

Ans. 7/8

Q.41 If three coins are tossed, what is the probability of getting fewer than two heads?

Ans. 1/2

Q.42 A coin is tossed three times and the outcome is recorded for each toss. Let A be the event that the experiment yields two tails. List the sample points in A.

Ans. {(H, T, T), (T, H, T), (T, T, H)}

Q.43 If we roll a balanced die, what is the probability of getting an even number?

Ans. 1/2

Q.44 Two dice are rolled. Count the total number of sample points in the sample space S.

Ans. 36

Q.45 A pair of dice is thrown. What is the probability that the sum equals 2?

Ans. 1/36

Q.46 A die is rolled two times. What is the probability that the sum of the numbers observed will be greater than 9?

Ans. 1/6

Q.47 A pair of dice is rolled. What is the probability that the sum equals 7?

Ans. 1/6

Q.48 Suppose two balanced dice are rolled. What is the probability that the sum of the dice is even?

Ans. 1/2

Q.49 What is the probability of rolling a eleven with a pair of balanced dice?

Ans. 1/18

Q.50 When a card is picked from a well-shuffled deck of 52 playing cards, what is the probability that it will be a black ace?

Ans. 1/26

Q.51 If a card is chosen from a well-shuffled deck of 52 playing cards, what is the probability of selecting a 10 of any suit?

Ans. 1/13

Q.52 A card is selected from a deck of playing cards. What is the probability that the card is a king or queen?

Ans. 2/13

Q.53 A card is selected from a deck of playing cards. What is the probability that the card is a red or black?

Ans. 1

Q.54 A card is drawn at random from a deck of 52 playing cards. What is the probability that it will be an ace or a face card?

Ans. 4/13

Q.55 A card is drawn at random from a deck of 52 playing cards. What is the probability that it will be a spade or a face card?

Ans. 11/26

Q.56 What is the probability of obtaining two kings in drawing two cards from a shuffled deck of cards? If the first card is not replaced before the second card is drawn.

Ans. 1/221

Q.57 Two cards are drawn at random from an ordinary deck of 52 cards. Determine the probability that both cards are aces, if the first card is replaced before the second card is drawn.

Ans. 1/169

Q.58 Consider drawing two cards from a pack of 52 playing cards:

Let Event A = the first card is a heart and

Event B = the second card is a red card

Calculate $P(A \text{ and } B)$.

Ans. $P(A) = 1/4$, $P(B/A) = 25/51$, $P(A \text{ and } B) = 0.123$

Q.59 Suppose $P(A) = 1/3$, $P(A \cup B) = 1/2$ and $P(A \cap B) = 1/10$. Find $P(B)$.

Ans. 0.27

Q.60 Suppose $P(A) = 1/4$, $P(B) = 1/3$ and $P(A \cup B) = 1/2$. Determine $P(A \cap B)$.

Ans. 1/12

Q.61 Given $P(A) = 1/3$ and $P(B) = 1/4$. Suppose A and B are independent events, determine $P(A \cup B)$.

Ans. 1/2

Q.62 If $P(A) = 0.2$, $P(B) = 0.4$ and $P(A/B) = 0.375$, find $P(A \text{ and } B)$.

Ans. 0.15

Q.63 Given $P(A/B) = 1/5$ and $P(B/A) = 1/2$. Find $P(A)/P(B)$

Ans. 2/5

Q.64 Given $P(A) = 1/3$, $P(B) = 1/4$ and $P(A/B) = 1/6$. Find $P(B/A)$.

Ans. 1/8

Q.65 If $P(A) = 0.7$, $P(B) = 0.5$ and $P(B/A) = 0.5$, find $P(A \text{ and } B)$.

Ans. 0.35

Q.66 If P(A and B) = 0.4 and P(B) = 0.8, find P(A/B).

Ans. 0.5

Q.67 Given P(A) = 0.7 and P(B) = 0.6. If A and B are statistically independent then find P(A and B).

Ans. 0.42

Q.68 Given P(A) = 0.3, P(B) = 0.4 and P(A and B) = 0.20. Are A and B statistically independent?

Ans. No, because P(A and B) $\neq$ P(A) P(B)

Q.69 A coin is tossed four times and the outcome is recorded for each toss. List the sample points for the experiment.

Ans. There are 16 equally likely outcomes:

HHHH	HHHT	HHTH	HHTT	HTHH	HTHT	HTTH	HTTT
THHH	THHT	THTH	THTT	TTHH	TTHT	TTTH	TTTT

Q.70 Given P(A) = 0.4, P(A $\cup$ B) = 0.7 and P(B) = p. Suppose A and B are independent events, find the value of **p**.

Ans. 0.5

Q.71 Two coins are tossed. Determine the probability of getting two heads given that there is at least one head.

Ans. 1/3

Q.72 Given P(A) = 2/3, P(A $\cup$B) = 7/12 and P(A$\cap$B) = 5/12. Suppose A and B are any two events, determine P(B).

Ans. 1/3

Q.73 Given P(A) = 1/3, P(B) = 3/4 and P(A$\cup$B) = 11/12. Find P(B/A).

Ans. 1/2

Q.74 A fair die is rolled. Find the probability of getting even face or face more than 4.

Ans. 2/3

Q.75 A fair die is rolled. What is the probability of getting even face given that face is less than 5?

Ans. 1/2

Q.76 Two dice are rolled. What is the probability of getting a total of 4 given that both faces are similar.

Ans. 1/6

Q.77 Given P(A) = 0.60 and P(B) = 0.10. If A and B are not mutually exclusive events, then find P(A $\cup$ B).

Ans. 0.64

EXERCISES

Q.1 Two coins are tossed once. Find the following probabilities:

(i) exactly two heads (ii) exactly two tails (iii) at least one head

(iv) at most two heads (v) at least one tail (vi) at most one tail

(vii) equal number of heads and tails (viii) no head

Ans: (i) 1/4 (ii) 1/4 (iii) 3/4 (iv) 1 (v) 3/4 (vi) 3/4 (vii) 1/2 (viii) 1/4

Q.2 Three coins are tossed. What is the probability of getting (i) exactly 2 heads, (ii) at most 2 heads (iii) at least 2 heads (iv) greater number of tails than heads, (v) equal number of heads and tails.

Ans. (i) 3/8 (ii) 7/8 (iii) 1/2 (iv) 1/2 (v) 0

Q.3 Two six sided dice are rolled. What is the probability that (i) the numbers appearing on both dice are same (ii) the sum of the numbers on two dice is 8 (iii) product of the numbers on two dice is 6.

Ans. (i) 1/6 (ii) 5/36 (iii) 1/9

Q.4 Playing with two balanced dice, what is the probability that (i) total on the two dice is 2 (ii) total is minimum or maximum (iii) both faces are alike (iv) total is a complete square (v) product of faces is a complete square (vi) there is at least one ace.

Ans. (i) 1/36 (ii) 1/18 (iii) 1/6 (iv) 7/36 (v) 2/9 (vi) 11/36

Q.5 Show that in a single throw with two dice, the chance of throwing more than 7 is equal to that of throwing less than 7.

Ans. 15/36, 15/36

Q.6 An income tax officer has received 10 files numbered from 1 to 10. He selects just one file for inspection. Find the probability that :

(i) the file number is multiple of 5 (ii) the file number is multiple of 3

(iii) the file number is multiple of 5 or multiple of 3.

Ans. (i) 1/5 (ii) 3/10 (iii) 1/2

Q.7 A card is drawn at random from an ordinary deck of playing cards. Find the probability that (i) card is red, (ii) card is an ace, (iii) card is of diamonds, (iv) card is 2, 3, 4, 5 or 6, (v) card is a red queen.

Ans. (i) 1/2 (ii) 1/13 (iii) 1/4 (iv) 5/13 (v) 1/26

Q.8 If three cards are chosen from a well-shuffled deck of 52 playing cards, what is the probability of selecting

(i) 3 hearts (ii) no heart (iii) 3 red cards (iv) no red card?

Ans: (i) 0.0129 (ii) 0.4135 (iii) 0.1176 (iv) 0.1176

Q.9 A card is selected at random from a well shuffled deck of 52 playing cards. Consider the following events:

Let A = Ace card B = Face card C = King card

(i) Show that A and B are mutually exclusive events

(ii) Show that A and C are equally likely events

(iii) Are B and C mutually exclusive events?

(iv) Are B and C equally likely events?

Ans: (i) Mutually exclusive (ii) Equally likely (iii) No (iv) No

Q.10 Let A denotes a card of heart, B a faced card. $\bar{A}$ and $\bar{B}$ are their complementary events. If a card is drawn at random from a full deck, find the following probabilities

(i) $P(A)$ (ii) $P(A \cap B)$ (iii) $P(A \cup B)$ (iv) $P(A \cap \bar{B})$ (v) $P(A \cup \bar{B})$

Ans. (i) 1/4 (ii) 3/52 (iii) 11/26 (iv) 5/26 (v) 43/52

Q.11 A coin is tossed twice. Let A be the event that at least one tail appears, B be the event that at most one tail appears and C be the event that exactly two tails appear.

(i) Are B and C mutually exclusive events?

(ii) Are A and B equally likely events?

(iii) Are B and C collectively exhaustive events?

Ans: (i) Yes (ii) Yes (iii) Yes

Q.12 A die is thrown. Consider the following events:

Let A = even number occurs B = prime number occurs

C = odd number occurs D = number is divisible by 5

(i) Are A and B equally likely events?

(ii) Are A and B mutually exclusive events?

(iii) Are A and C mutually exclusive events?

(iv) Are A and C collectively exhaustive events?

(v) Which of the above events are simple and which are compound?

Ans: (i) Yes (ii) No (iii) Yes (iv) Yes

(v) Simple event = D and compound events = A, B and C.

Q.13 Two dice are thrown once. The following events are defined:

A = sum of spots is 6 B = sum of spots is 8

C = same numbers appear D = different numbers appear

(i) Are A and B mutually exclusive events?

(ii) Are A and C mutually exclusive events?

(iii) Are A and B equally likely events?

(iv) Are C and D collectively exhaustive events?

Ans: (i) Yes (ii) No (iii) Yes (iv) Yes

Q.14 A bag contains 4 white and 2 black balls. Two balls are selected at random without replacement. Find the following probabilities after making the sample space of all possible outcomes.

(i) both are white (ii) both have the same colour (iii) both have different colours.

Ans. (i) 2/5 (ii) 7/15 (iii) 8/15

Q.15 A bag contains 7 white, 5 black and 4 red balls. If two balls are drawn at random, find the probability that (i) both the balls are white (ii) one is black and other is red.

Ans. (i) 7/40 (ii) 1/6

Q.16 A committee of 4 people is to be appointed from 3 officers of the production department, 4 officers of the purchase department, 2 officers of the sales department and 1 chartered accountant. Find the probability of forming the committee in the following manner:

(i) There must be one from each category.

(ii) It should have at least one from the purchase department.

(iii) The chartered accountant must be in the committee.

Ans. (i) 24/210 (ii) 13/14 (iii) 2/5

Q.17 An urn contains 8 red, 3 white and 2 blue balls. If 3 balls are drawn at random, determine the probability that (i) all 3 are red (ii) all 3 are white (iii) 2 are red and 1 is blue (iv) at least one is white (v) the balls are drawn in the order, red, white and blue.

Ans. (i) 28/143 (ii) 1/286 (iii) 28/143 (iv) 83/143 (v) 4/143

Q.18 Six cards are drawn at random from a deck of 52 cards. What is the probability that 3 are red and 3 are black?

Ans. 0.332

Q.19 If the probability of a horse A winning a race is 1/5 and that of a horse B is 1/6, what is the probability that one of them wins?

Ans. 11/30

Q.20 A coin and a die are thrown together. Find the chance of throwing a head and 5 or a tail and 6.

Ans. 1/6

Q.21 What is the probability of throwing either 7 or more than 10 with two dice?

Ans. 1/4

Q.22 Two fair coins are tossed. Let A denote the event "both faces are heads" and B denote the event "only one face is head". Are A and B disjoint events. Evaluate $P(A \cup B)$.

Ans. Yes, 3/4

Q.23 An integer is chosen at random from the first 200 positive integers. What is the probability that the integer chosen is divisible by 6 or 8?

Ans. 1/4

Q.24 The integers 1, 2, 3, ⋯, 20 are written on slips of paper which are placed in a bowl and thoroughly mixed. A slip is drawn. What is the probability that the number on the slip is:

(i) either square or divisible by 3? (ii) either prime or odd number?

Ans. (i) 9/20 (ii) 11/20

Q.25 An integer is chosen at random from the first 100 positive integers. Find the probability that the chosen digit is

(i) multiple of 10 (ii) divisible by 8 (iii) divisible by 8 or 12.

Ans. (i) 1/10 (ii) 3/25 (iii) 4/25

Q.26 A pair of dice is thrown, find the probability that (i) the sum of the digits on turned up faces is either 6 or 9 and (ii) it is either a doublet or less than 5.

Ans. (i) 1/4 (ii) 5/18

Q.27 If A and B be events with $P(A \cup B) = 7/8$, $P(A \cap B) = 1/4$ and $P(\bar{A}) = 5/8$.

Find $P(A)$, $P(B)$ and $P(A \cap \bar{B})$.

Ans. $P(A) = 3/8$, $P(B) = 3/4$, $P(A \cap \bar{B}) = 1/8$

Q.28 The following contingency table is set up:

	B	$\bar{B}$
A	10	20
$\bar{A}$	20	40

Compute the following probabilities

(i) event A (ii) event B (iii) event $\bar{A}$

(iv) event $\bar{B}$ (v) event A and B (vi) event $\bar{A}$ and B

(vii) event A and $\bar{B}$ (viii) event $\bar{A}$ and $\bar{B}$ (ix) event A or B

(x) event $\bar{A}$ or B (. i) event A or $\bar{B}$ (xii) event $\bar{A}$ or $\bar{B}$

Ans: (i) 1/3 (ii) 1/3 (iii) 2/3 (iv) 2/3 (v) 1/9 (vi) 2/9

(vii) 2/9 (viii) 4/9 (ix) 5/9 (x) 7/9 (xi) 7/9 (xii) 8/9

Q.29 Two coins are tossed. What is the conditional probability that the two heads result, given that there is at least one head?

Ans. 1/3

Q.30 A fair die is rolled. Let A be the event that the face is prime and B be the event that the face is even. Find the following probabilities.

(i) $P(A \cup B)$ (ii) $P(A \cap B)$ (iii) $P(A/B)$ (iv) $P(B/A)$ (v) $P(A/\bar{B})$

Ans. (i) 5/6 (ii) 1/6 (iii) 1/3 (iv) 1/3 (v) 2/3

Q.31 A pair of dice is thrown. If the two numbers appearing are different, find the probability that (i) the sum is 6 (ii) the sum is four or less.

Ans. (i) 2/15 (ii) 2/15

Q.32 A card is taken at random from playing cards, find the probability that card is a king, when it is given (known) that the card is

(i) red (ii) of diamonds (iii) faced.

Ans. (i) 1/13 (ii) 1/13 (iii) 1/3

Q.33 In a firm, 20 percent of the employees have accounting background, while 5 percent of the employees are executives and have accounting background. If an employee has an accounting background, what is the probability that the employee is an executive?

Ans. 1/4

Q.34 The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$, the probability that it arrives on time is $P(A) = 0.92$, and the probability that it departs and arrives in time is $P(D \cap A) = 0.78$. Find the probability that a plane (i) arrives in time given that it departed on time, and (ii) departed on time given that it has arrived in time.

Ans. (i) 0.94 (ii) 0.85

Q.35 Given the following contingency table:

	B	$\bar{B}$
A	20	40
$\bar{A}$	10	30

What is the probability of

(i) A / B? (ii) $A / \bar{B}$? (iii) $\bar{A} / B$? (iv) $\bar{A} / \bar{B}$?

Ans: (i) 2/3 (ii) 4/7 (iii) 1/3 (iv) 3/7

Q.36 Two dice are rolled. Let E_1 denote the event of an odd total, E_2 the event of an ace on the first die and E_3 the event of a total seven.

(i) Are E_1 and E_2 independent? (ii) Are E_1 and E_3 independent?

(iii) Are E_2 and E_3 independent?

Ans. (i) independent (ii) dependent (iii) independent.

Q.37 The probability that a boy will pass an examination is 3/5 and that for a girl it is 2/5. What is the probability that at least one of them will pass the examination?

Ans. 19/25

Q.38 Two urns contain respectively 3 white, 7 red, 15 black and 10 white, 6 red and 9 black balls. One ball is drawn from each urn. What is the probability that both balls are of the same colour.

Ans. 0.3312

Q.39 The probability that a 50 year old man will be alive at 60 is 0.83 and the probability that a 45 year old woman will be alive at 55 is 0.87. What is the probability that a man who is 50 and his wife who is 45 will be alive 10 years hence? Also find the probability that at least one of them will be alive 10 years hence.

Ans. 0.7221, 0.9779

Q.40 Three missiles are fired at a target. If the probabilities of hitting the target are 0.4, 0.5 and 0.6, respectively and if the missiles are fired independently, what is the probability that at least two missiles will hit the target?

Ans. 0.5

Q.41 A can hit a target 4 times in 5 shots, B can hit 3 times in 4 shots and C can hit twice in 3 shots. They fire a volley. What is the probability that the two shots at least hit?

Ans. 5/6

Q.42 A husband and a wife appear in an interview for two vacancies of the same post. The probability of husband's selection is 1/7 and that of wife's selection is 1/5. What is the probability that (i) both of them will be selected (ii) only one of them will be selected (iii) none of them will be selected.

Ans. (i) 1/35 (ii) 10/35 (iii) 24/35

Q.43 A and B play 12 games of chess out of which 6 are won by A, 4 are won by B, and 2 end in a tie. They agree to play a tournament consisting of 3 games. Find the probability that (i) A wins all three games (ii) two games end in a tie (iii) A and B win alternately (iv) B wins at least one game.

Ans. (i) 1/8 (ii) 5/72 (iii) 5/36 (iv) 19/27

Q.44 Two cards are drawn from a well-shuffled ordinary deck of 52 cards. Find the probability that they are both aces if the first card is

(i) replaced (ii) not replaced.

Ans. (i) 1/169 (ii) 1/221

Q.45 Three balls are drawn successively from a box containing 6 red, 4 white and 5 blue balls. Find the probability that they are drawn in the order, red, white and blue if each ball is (i) replaced (ii) not replaced.

Ans. (i) 8/225 (ii) 4/91

Q.46 Find the chance of drawing an ace, a king, a queen and a jack in order from an ordinary pack in four consecutive draws, the cards drawn being not replaced.

Ans. $\frac{32}{812175}$

Q.47 Urn A contains five red and three white balls. Urn B contains two red and six white balls. (i) If a ball is drawn from each urn, what is the probability that both are of the same colour. (ii) If two balls are drawn from each urn, what is the probability that all the four balls are of the same colour?

Ans. (i) 7/16 (ii) 55/784

Q.48 If A and B are independent events and $P(A) = 0.30$ and $P(B) = 0.60$, find

(i) $P(A/B)$ (ii) $P(A \cap B)$ (iii) $P(A \cup B)$ (iv) $P(\bar{A} \cap \bar{B})$

Ans: (i) 0.30 (ii) 0.18 (iii) 0.72 (iv) 0.28

Q.49 Given $P(A) = 0.60$, $P(B) = 0.40$, $P(A \cap B) = 0.24$. Find $P(A/B)$, $P(B/A)$ and $P(A \cup B)$. What is the relationship between A and B?

Ans. $P(A/B) = 0.6$, $P(B/A) = 0.4$, $P(A \cup B) = 0.76$, A and B are independent

Q.50 Assume that events A and B are independent. That is $P(A/B) = P(A)$ and $P(B/A) = P(B)$. You are given the probability assessments that $P(A) = 0.45$ and $P(B) = 0.80$. Find: (i) $P(A \cap B)$ (ii) $P(A \cup B)$.

Ans. (i) 0.36 (ii) 0.89

Q.51 A bag contains 5 blue and 3 black balls and another bag contains 3 blue and 5 black balls. One bag is selected at random and a ball is taken out of it. Find the probability that ball is black.

Ans. 1/2

Q.52 A problem in statistics is given to three students A, B and C whose chances of solving it are 1/3, 1/4 and 1/2 respectively. What is the probability that the problem will be solved?

Ans. 3/4

Q.53 A, B, C, D cut a pack of cards successively in the order mentioned. What are their respective chances of cutting a spade first ?

Ans. $P(A) = 64/175$, $P(B) = 48/175$, $P(C) = 36/175$, $P(D) = 27/175$

Q.54 A, B and C in order toss a coin. The first one to throw a head wins. What are their respective chances of winning? Assume that the game may continue indefinitely.

Ans. $P(A) = 4/7$, $P(B) = 2/7$, $P(C) = 1/7$

Chapter

8

RANDOM VARIABLE AND PROBABILITY DISTRIBUTIONS

8.1. INTRODUCTION

The word random is used quite commonly in our daily life. One may or may not know its meaning but whenever it is used, it conveys the sense for which it is used. The customers enter a shop not according to some preplanned manner, but they enter the shop in a random manner. The vehicles cross a zebra crossing in a random manner. The teacher does not check the note-books of all students, he checks some of the note-books selected at random. Somebody is intelligent by birth, somebody is healthy by birth, somebody faces an accident on the road, somebody gets an attack of influenza. All these events occur randomly. When ever a coin is tossed or a die is thrown, the results happen in a random manner. Many situations in practical life are of random nature. Their ultimate results are based on chance. Even a young kid is also familiar with the classical idea of "lottery method", which is centuries old and has been used for the selection of a random sample. Most modern methods of selecting a sample are based on the theory of random selection by lottery. The random sample is the basis of the statistical inference. Thus randomness is the central idea of the study which is carried out to know something about unknown situations.

8.2. GENERATION OF RANDOM NUMBERS

In our counting system, there are ten basic digits which are used for counting purposes. These digits are 0, 1, 2, ... 9. We can make numbers of any size with the help of these digits. The figure 53792 is made up of five digits 2, 3, 5, 7 and 9. We shall use these digits to make a set of numbers called table of random numbers. Suppose we select ten paper slips and on each slip we write a different digit. Thus each slip represents a digit. We select any one of these slips at random and note down its digit on a paper. We return the slip to the main lot and select a slip again. The digit on the second slip is also noted along with the first digit (row-wise) or below the first digit (column-wise). We continue this process of selecting, recording and replacing each selected slip. On each selection the probability of selection of each digit is 1/10. Thus each digit has equal probability of selection. We get a set of

digits called random digits. If the first digit is 5, second is 7, third is 5 and fourth is 0, we can write them in a row as 5750 or 57 50. We can also write in a column as below:

5
7
5
0

When the first row or column is completed, we can write the selected digits in the second row or second column. In this manner a table of any size spread over a number of pages can be obtained. This is called table of random numbers. One small table of random numbers is given below:

Table 8.1.

5 1	2 2	0 9	1 2	7 2	1 2	4 0	9 2
7 2	4 5	3 5	5 0	2 3	3 9	7 4	4 4
5 7	1 8	7 3	3 1	1 1	7 5	8 8	7 5
9 2	6 9	4 6	7 5	5 6	8 2	7 7	6 6
3 8	3 2	1 2	9 3	9 5	6 8	8 4	8 7
9 5	7 1	8 0	3 6	8 2	1 6	4 8	3 8

This table is written with two digits in two columns together. This is one way of writing the digits. One can write 3-digit or 4-digit columns. Anybody can make a table of random numbers. A good table of random numbers contains 0, 1, 2, ..., 9 almost equal number of times. The students shall learn in higher classes that the random numbers can be made for each probability distribution. The random numbers under discussion, infact are the random numbers from a discrete uniform distribution over the interval (0, 9).

8.3. HISTORY OF RANDOM NUMBER TABLES

In actual practice, we do not have to make the table of random numbers. These tables are given in the books and we can consult them. The first table of random numbers was constructed by Tippett in 1927. The digits in the logarithms of various integers were used to make this table. Fisher and Yates prepared a table of 15000 numbers. Another table of random numbers was constructed by Smith and Kendall in 1939. This was prepared by revolving a wheel on a machine. The wheel was stopped at random. The wheel was divided into ten equal parts and each part indicated a digit. In 1955, the Rand Corporation published a book "A million random numbers". One page of this random number table is given in end of this book. The Rand Corporation prepared the random numbers with the help of computer. We can also get the random numbers from the electronic calculators which have the function 'Ran' (for random) on them.

8.4. APPLICATION OF RANDOM NUMBERS

There are many uses of random numbers in statistics. One of the important use is in the selection of a simple random sample from a finite population. Suppose there are 80 students in a class and we want to select 8 students at random. The students can be numbered from 01 to 80. Then we consult the random number table. Let us see Table 8.1. We shall read two-digit columns. If any number is between 01 and 80, we shall note it down for the sample. If any random number is above 80, we shall ignore it because it is not in our population. We can read the random number table from any place. We may move column-wise or row-wise. Let us read the random numbers from Table 8.1. We read the first column. The random numbers are 51, 72, 57, 38, 22, 45, 18, 69. Two random numbers 92 and 95 have been ignored because they are above 80. The remaining eight random numbers represent those eight students who have been selected for the sample. If the number of units in the population are in hundreds, say 800 we shall read three-digit column of the random number table. If we have 100 units in the population, we shall number them from 00 to 99 so that we have to read 2-digit columns. If we number them as 001, 002, ... 100, we shall have to read 3-digit columns. In three digit column, most of the random numbers will be above 100 and a lot of time will be wasted in getting the required size of the sample. Further use of random number table for the selection of a simple random sample will be discussed in a chapter on sampling and sampling distributions.

Random number table can also be used to generate data without performing the actual experiment. Suppose we want to toss a coin 10 times to see the number of heads. We can do it by

(i) tossing the coin and counting the number of heads.

(ii) By using random number table. The even digits 0, 2, 4, 6, 8 will stand for the head and the odd digits 1, 3, 5, 7, 9 will be for the tail. The first ten digits of the first row of Table 8.1. are reproduced below. H is for head and T is for tail

Digits:	5	1	2	2	0	9	1	2	7	2
Outcome:	T	T	H	H	H	T	T	H	T	H

There are 5 heads and 5 tails. It is just a chance that number of heads is equal to the number of tails. If we read the next row, the result in general would be different.

This process of getting the results of an experiment from random number table is called simulation. In simulation the actual experiment is not performed.

Example 8.1.

Two balanced coins are to be tossed 10 times to record the number of heads each time. Use random number table to record the possible observations. Write the frequency distribution of the observed number of heads.

Solution:

Two digits of a random number table will represent the result of a throw of two coins. We shall take ten pairs of random numbers for 10 throws of two coins.

From Table 8.1. we take 10 pairs of random digits and count the number of heads. Even digit will indicate head (H) and an odd digit will indicate tail (T)

Random pairs:	51	22	09	12	72	12	40	92	72	95
Number of even digits	0	2	1	1	1	1	2	1	1	0
No. of heads:	0	2	1	1	1	1	2	1	1	0

The observed frequency distribution of number of heads is

Number of heads	Frequency
0	2
1	6
2	2
Total	10

If two coins are actually tossed 10 times, the observed frequency distribution may or may not agree with the above distribution.

Note: If 3 coins are to be tossed, we shall take 3 digits to represent a throw of 3 coins.

Example 8.2.

Assume that a fair die is to be rolled ten times. Without rolling the die, obtain the possible outcomes using random digits.

Solution:

Here the six digits 1, 2, 3, 4, 5, 6 will represent the six faces of the die. We shall ignore the random digit 0 and the digits above 6.

From Table 8.1. we have the random digits

5 1 2 2 1 2 2 1 2 4

Thus we assume that the first throw has given 5 on the die, 2nd throw has given 1 on the die and the 10th throw has given 4 on the die.

Example 8.3.

Assume that a die is to be tossed ten times. Use random digits to find the number of successes where success is attained through 1.

Solution:

From Table 8.1. we take the random digits as below. The digit 1 is for success (S) and digits 2, 3, 4, 5, 6 are for failure (F) and the digit 0 and the digits above 6 will be ignored.

Random digits:	5	1	2	2	1	2	2	1	2	4
Result:	F	S	F	F	S	F	F	S	F	F

Thus in 10 throws, the success has been achieved 3 times.

8.5. RANDOM VARIABLE

A set of numerical values assigned to the all possible outcomes of the random experiment is called random variable. The random variable can be briefly written as r.v. If we write A, B, ... , F on the six faces of a die, these letters are not a r.v. If we

write some numerical values on the six faces of a die like 1, 2, 3, ... , 6, we have a set of values called r.v. Suppose we select two bulbs from a certain lot having good and defective bulbs. Let G stand for good and D stand for defective. There are four possible outcomes which are GG, GD, DG and DD. Each outcome can be assigned some numerical value. Let us count number of defective bulbs in each outcome. We can write

Outcome	No. of defective bulbs
GG	0
GD	1
DG	1
DD	2

Thus the numerical values 0, 1, 2 are the values of the random variable where random variable is the number of defective bulbs in this discussion. A random variable is denoted by a capital letter X. Here X is the number of defective bulbs. The small letters $x_1, x_2, \dots, x_n$ are used for the specific values of the random variable. A random variable is also called chance variable. If we have two or more than two random variables we can use the letters X, Y, Z for them. A random variable may be discrete or continuous.

8.6. DISCRETE RANDOM VARIABLE

A random variable X is called discrete if it can assume finite or countably infinite number of values. If two bulbs are selected from a certain lot, the number of defective bulbs may be 0, 1 or 2. The range of the variable is from 0 to 2 and random variable can take some selected values in this range. The number of defective bulbs cannot be 1.1 or −1 or 3 etc. This random variable can take only the specific values which are 0, 1 and 2. When two dice are rolled the total on the two dice will be 2, 3, ... , 12. The total on the two dice is a discrete random variable.

8.7. DISCRETE PROBABILITY DISTRIBUTION

Suppose a discrete random variable X can assume the values $x_1, x_2, x_3, \dots, x_n$ with corresponding probabilities $p(x_1), p(x_2), p(x_3), \dots, p(x_n)$. The set of ordered pairs $[x_1, p(x_1)], [x_2, p(x_2)], [x_3, p(x_3)], \dots, [x_n, p(x_n)]$ is called the probability distribution or probability function of random variable X. A probability distribution can be presented in a tabular form showing the values of the random variable X and the corresponding probabilities denoted by $p(x_i)$ or $f(x_i)$. The above information can be collected in the form of a table below called the probability distribution of the random variable X. The probability that random variable X will take the value x_i is denoted by $p(x_i)$ where $p(x_i) = P(X = x_i)$

Values of random variable (x_i)	x_1	x_2	x_3	...	x_i	...	x_n
Probability $p(x_i)$	$p(x_1)$	$p(x_2)$	$p(x_3)$	...	$p(x_i)$	...	$p(x_n)$

The probability of some interval can be calculated by adding the probabilities of all the points in the interval. For example $P[x_1 < X < x_4] = P(X = x_2) + P(X = x_3)$.

This type of addition will not be possible in continuous random variable for finding the probability of an interval. Therein we shall use the integral calculus for finding the probability of an interval.

The probability distribution can also be described in the form of an equation for $f(x_i)$ with a list of possible values of the random variable X. Some probability distributions in the form of equations are

(i) $f(x_i) = \frac{1}{6}$ for $x_i = 1, 2, 3, \ldots, 6.$

For each value of X the probability is 1/6. It is the probability distribution when a fair die is rolled.

(ii) $f(x_i) = \binom{3}{x}\left(\frac{1}{2}\right)^3$ for $x = 0, 1, 2, 3.$

It is the probability distribution when three fair coins are tossed.

8.8. PROPERTIES OF THE DISCRETE PROBABILITY FUNCTION

A discrete probability distribution or function $f(x_i)$ has the following properties.

(i) f(x) is always non-negative. It lies between 0 and 1. Thus $0 \le f(x) \le 1$

(ii) The sum of all the probabilities is 1. Thus $\sum_{i=1}^{n} f(x_i) = 1$

8.9. GRAPH OF THE DISCRETE PROBABILITY FUNCTION

The discrete probability distribution can be presented in the form of a graph by taking the values $x_1, x_2, x_3, \ldots, x_i, \ldots$ on X-axis and drawing vertical lines of heights equal to $f(x_1), f(x_2), f(x_3), \ldots, f(x_i), \ldots$ above them. This type of graph is called line graph.

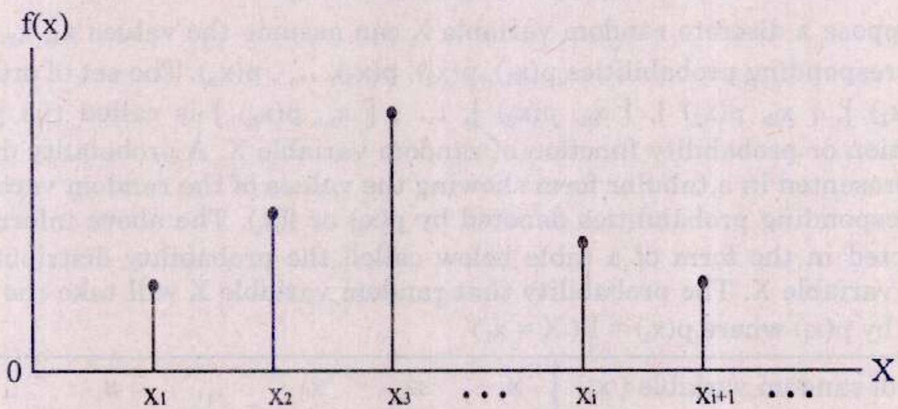


Fig.8.1. Line graph of the probability distribution.

8.9.1. PROBABILITY HISTOGRAM

The probability distribution can also be presented in the form of probability histogram.

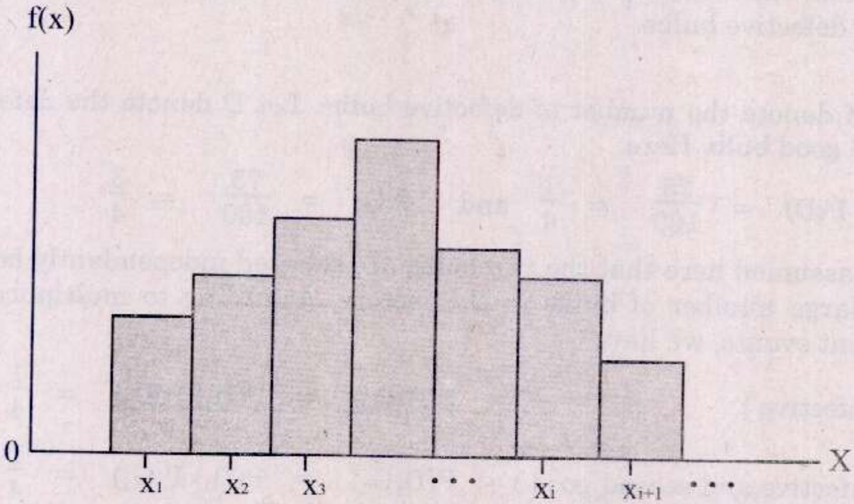


Fig.8.2. Probability histogram

Example 8.4.

A digit is selected from the first 8 natural numbers. Write the probability distribution of X where X is the number of factors (divisors) of the digits.

Solution:

It is assumed here that the probability of selection for each digit is 1/8. We can write:

Digit	Factors	Number of factors	Probability
1	1	1	1/8
2	1, 2	2	1/8
3	1, 3	2	1/8
4	1, 2, 4	3	1/8
5	1, 5	2	1/8
6	1, 2, 3, 6	4	1/8
7	1, 7	2	1/8
8	1, 2, 4, 8	4	1/8

The information in the above table can be summarised as below in the form of table of probability distribution.

Number of factors r.v. (x_i)	1	2	3	4	Total
p(x_i)	1/8	4/8	1/8	2/8	1

The probability of X = 2 is 4/8 which has been obtained by adding 1/8 four times. Similarly the probability of 4 is 2/8 which has been obtained by adding 1/8 two times.

Example 8.5.

A factory is producing bulbs out of which 25 % are defective. Two bulbs are selected from this factory for inspection. Write the probability distribution of number of defective bulbs.

Solution:

Let X denote the number of defective bulbs. Let D denote the defective and G denote the good bulb. Here

$$P(D) = \frac{25}{100} = \frac{1}{4} \quad \text{and} \quad P(G) = \frac{75}{100} = \frac{3}{4}$$

It is assumed here that the two bulbs are selected independently because there are very large number of bulbs in the factory. According to multiplication law of independent events, we have

$$\text{P(both defective)} = P(D_1D_2) = P(D_1)\,P(D_2) = \frac{1}{4}\times\frac{1}{4} = \frac{1}{16}$$

$$\text{P(first defective and second good)} = P(D_1G_2) = P(D_1)\,P(G_2) = \frac{1}{4}\times\frac{3}{4} = \frac{3}{16}$$

$$\text{P(first good and second defective)} = P(G_1D_2) = P(G_1)\,P(D_2) = \frac{3}{4}\times\frac{1}{4} = \frac{3}{16}$$

$$\text{P(both good)} = P(G_1G_2) = P(G_1)\,P(G_2) = \frac{3}{4}\times\frac{3}{4} = \frac{9}{16}$$

The possible outcomes, the random variable X and the corresponding probabilities are put in the following table :

Outcomes (S)	Number of defectives (r.v.)	Probability
D_1D_2	2	$\frac{1}{4}\times\frac{1}{4} = \frac{1}{16}$
D_1G_2	1	$\frac{1}{4}\times\frac{3}{4} = \frac{3}{16}$
G_1D_2	1	$\frac{3}{4}\times\frac{1}{4} = \frac{3}{16}$
G_1G_2	0	$\frac{3}{4}\times\frac{3}{4} = \frac{9}{16}$

The above information can be collected as below in the form of a probability distribution of the random variable X.

Random Variable (x_i)	0	1	2
$p(x_i)$	9/16	6/16	1/16

Example 8.6.

From an urn containing 4 red and 6 white round marbles, a man draws three marbles at random without replacement. If X is a random variable which denotes the number of red marbles drawn, what is the probability distribution of X? Draw a probability histogram.

Solution:

Red marbles	White marbles	Total marbles
4	6	10

S contains $\binom{10}{3} = 120$ sample points.

If the random variable X denotes the number of red marbles, the possible values of x are 0, 1, 2, and 3, and their respective probabilities are:

$$P(X=0) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{5}{30} \qquad P(X=1) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{15}{30}$$

$$P(X=2) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{9}{30} \qquad P(X=3) = \frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{1}{30}$$

The probability distribution of red marbles in a tabular form is:

x	0	1	2	3
p(x)	5/30	15/30	9/30	1/30

The probability distribution is shown as a histogram in Fig.8.3.

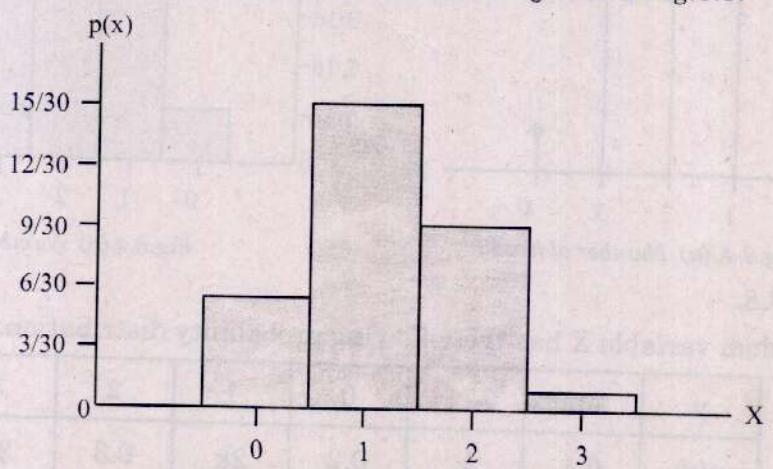


Fig.8.3. Number of red marbles.

Example 8.7.

Given the discrete probability distribution:

$$p(x) = \binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x} \quad \text{for } x = 0, 1, 2, 3, 4.$$

Find the complete probability distribution and draw suitable graphs.

Solution:

$$p(x) = \binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x} \quad x = 0, 1, 2, 3, 4$$

$$P(X=0) = \binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = \frac{1}{16} \qquad P(X=1) = \binom{4}{1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = \frac{4}{16}$$

$$P(X=2) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16} \qquad P(X=3) = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = \frac{4}{16}$$

$$P(X=4) = \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$$

The probability distribution of X in a tabular form is:

x	0	1	2	3	4
p(x)	1/16	4/16	6/16	4/16	1/16

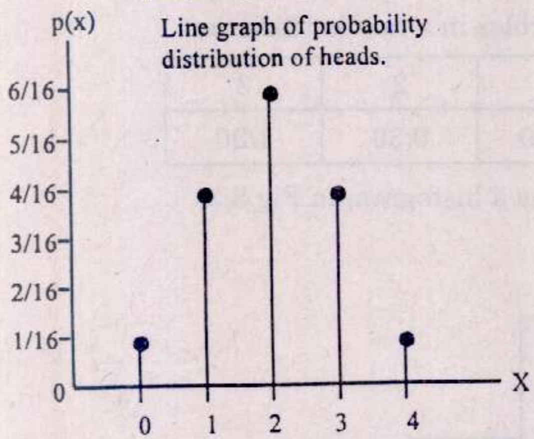


Fig.8.4. (a) Number of heads

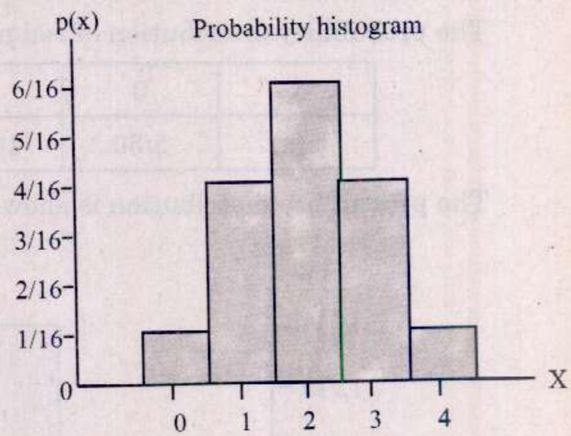


Fig.8.4.(b) Number of heads

Example 8.8.

A random variable X has the following probability distribution:

x	−2	−1	0	1	2	3
p(x)	0.1	k	0.2	2k	0.3	3k

Find: (i) k (ii) $P(X < 2)$ (iii) $P(X \geq 2)$ (iv) $P(-2 < X < 2)$ (v) $P(X \leq 1)$.

Solution:

(i) Since the sum of probabilities is one

$\sum_{x=-2}^{3} p(x) = 1$ gives

$0.1 + k + 0.2 + 2k + 0.3 + 3k = 1$ or $0.6 + 6k = 1$ or $6k = 1 - 0.6 = 0.4$

so that $k = \frac{0.4}{6} = \frac{4}{60} = \frac{1}{15}$

(ii) $P(X < 2) = P(X = -2) + P(X = -1) + P(X = 0) + P(X = 1)$

$= 0.1 + k + 0.2 + 2k = 0.3 + 3k = 0.3 + 3\left(\frac{1}{15}\right) = 0.3 + 0.2 = 0.5$

(iii) $P(X \geq 2) = P(X = 2) + P(X = 3) = 0.3 + 3k$

$= 0.3 + 3\left(\frac{1}{15}\right) = 0.3 + 0.2 = 0.5$

(iv) $P(-2 < X < 2) = P(X = -1) + P(X = 0) + P(X = 1) = k + 0.2 + 2k = 0.2 + 3k$

$= 0.2 + 3\left(\frac{1}{15}\right) = 0.2 + 0.2 = 0.4$

(v) $P(X \leq 1) = P(X = -2) + P(X = -1) + P(X = 0) + P(X = 1)$

$= 0.1 + k + 0.2 + 2k = 0.3 + 3k = 0.3 + 3\left(\frac{1}{15}\right) = 0.3 + 0.2 = 0.5$

8.10. CONTINUOUS RANDOM VARIABLE

A random variable is called continuous if it can assume all possible values in the possible range of the random variable. Suppose the temperature in a certain city in the month of June in the past many years has always been between 35° to 45° centigrade. The temperature can take any value between the range 35° to 45°. The temperature on any day may be 40.15°C or 40.16°C or it may take any value between 40.15° and 40.16°. When we say that the temperature is 40°C, it means that the temperature lies somewhere between 39.5° to 40.5°. Any observation which is taken falls in an interval. There is nothing like an exact observation in continuous variable. In discrete random variable the values of the variable are exact like 0, 1, 2 good bulbs. In continuous random variable the value of the variable is never an exact point. It is always in the form of an interval, the interval may be very very small.

Some examples of the continuous random variable are

(i) The computer time (in seconds) required to process a certain program

(ii) The time that a poultry bird will gain the weight of 1.5 kg.

(iii) The amount of rain fall in a certain city.

(iv) The amount of water passing through a pipe connected with a high level reservoir.

(v) The heat gained by a ceiling fan when it has worked for one hour.

8.11. PROBABILITY DENSITY FUNCTION

The probability function of the continuous random variable is called probability density function or briefly p.d.f. It is denoted by f(x) where f(x) is the probability that the random variable X takes the value between x and $x + \Delta x$ where Δx is a very very small change in X.

If there are two points 'a' and 'b' then the probability that the random variable will take the value between a and b is given by the integral

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

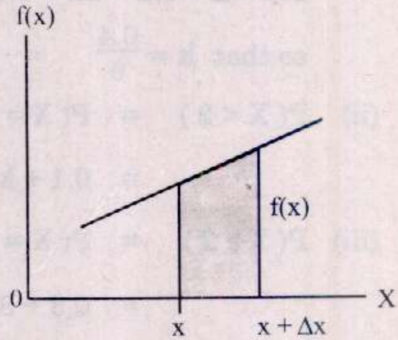


Fig.8.5.

where 'a' and 'b' are any points between $-\infty$ and $+\infty$. The quantity f(x) dx is called probability differential.

The number of possible outcomes of a continuous random variable is uncountably infinite. Therefore, a probability of zero is assigned to each point of the random variable. Thus $P(X = x) = 0$ for all values of X. This means that we must calculate a probability for a continuous random variable over an interval and not for any particular point. This probability can be interpreted as an area under the graph between the interval from a to b. When we say that the probability is zero that a continuous random variable assumes a specific value, we do not necessarily mean that a particular value cannot occur. We, in fact, mean that the point (event) is one of an infinite number of possible outcomes. Whenever we have to find the probability of some interval of the continuous random variable, we can use any one of these two methods.

(i) Integral calculus

(ii) Area by geometrical diagrams (this method is easy to apply when f(x) is a simple linear function).

8.12. PROPERTIES OF PROBABILITY DENSITY FUNCTION

The probability density function f(x) must have the following properties.

(i) It is non-negative i.e. $f(x) \geq 0$ for all x.

(ii) Total Area $= \int_{-\infty}^{\infty} f(x)\,dx = 1$

(iii) $P(X = c) = \int_c^c f(x)\,dx = 0$, where c is any constant.

(iv) As the probability is zero for X = c (constant), therefore $P(X = a) = P(X = b) = 0$. If we take an interval a to b, it makes no difference whether end points of the interval are considered or not. Thus we can write:

$$P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b)$$

(v) $$P(a \le X \le b) = \int_a^b f(x)\,dx = \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx \qquad (a < b)$$

Example 8.9.

Given a function $f(x) = c\,x, \quad 0 \le x \le 2$

(a) Find the value of c so that f(x) is a probability density function.

(b) Find the probabilities (i) $P(X < 1)$ (ii) $P\left(\frac{1}{2} < X < \frac{3}{2}\right)$

Solution:

a) f(x) will be a proper p.d.f. if the total area under the line and X-axis from 0 to 2 is unity. The graph of f(x) is a right angled triangle. Its area is given by

$$\text{Area} = \frac{\text{height} \times \text{base}}{2}$$

height $= f(x) = c\,x = 2c$ at $X = 2$

base $= 2 - 0 = 2$

Total Area has to be unity, therefore

$$\text{Area} = \frac{2c \times 2}{2} = 1$$

$$2c = 1, \text{ so that } c = \frac{1}{2}$$

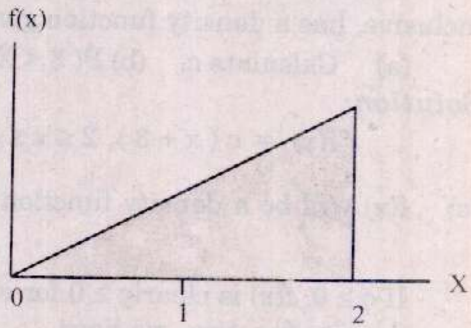


Fig. 8.6.(a)

Thus $f(x) = \frac{1}{2}x \quad 0 \le x \le 2$ is proper probability density function.

(b) (i) $P(X < 1)$ is the shaded area of the right-angled triangle.

$$\text{Area} = \frac{\text{height} \times \text{base}}{2}$$

$$f(x) = \frac{1}{2}x$$

height $(x = 1) = 1/2$

base $= 1 - 0 = 1$

$$\text{Area} = \frac{1/2 \times 1}{2} = \frac{1}{4}$$

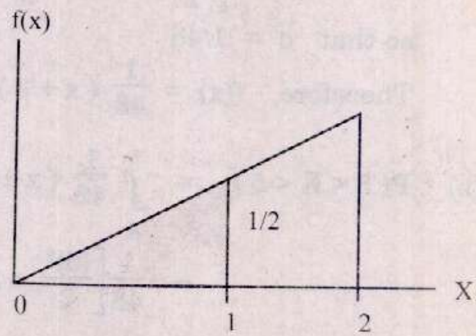


Fig. 8.6.(b)

(ii) The probability $P\left(\frac{1}{2} < X < \frac{3}{2}\right)$ is the area of the shaded portion of the diagram. The shaded area is a trapezoid and its area is given by

Area of Trapezoid = (Average height) (Base)

$$\text{Average Height} = \frac{\text{Sum of parallel sides}}{2}$$

$$= \frac{f\left(\frac{3}{2}\right) + f\left(\frac{1}{2}\right)}{2}$$

$$= \frac{3/4 + 1/4}{2} = \frac{1}{2}$$

$$\text{Base} = \frac{3}{2} - \frac{1}{2} = 1$$

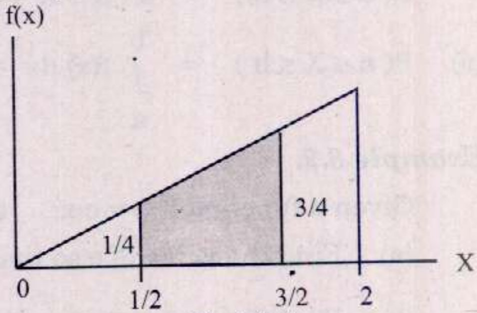


Fig.8.6.(c)

$$P\left(\frac{1}{2} < X < \frac{3}{2}\right) = \text{Area of Trapezoid} = 1/2 \times 1 = 1/2$$

Example 8.10.

A continuous random variable X which can assume values between x = 2 and 8 inclusive, has a density function given by c (x + 3) where c is a constant.

(a) Calculate c. (b) $P(3 < X < 5)$ (c) $P(X \geq 4)$.

Solution:

$$f(x) = c (x + 3),\ 2 \leq x \leq 8$$

(a) f(x) will be a density function if (i) $f(x) \geq 0$ for every x and (ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

If $c \geq 0$, f(x) is clearly ≥ 0 for every x in the given interval. Hence for f(x) to be a density function, we have

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{2}^{8} c(x+3)\,dx = c\left[\frac{x^2}{2} + 3x\right]_2^8$$

$$= c\left[\frac{(8)^2}{2} + 3(8) - \frac{(2)^2}{2} - 3(2)\right] = c[32 + 24 - 2 - 6] = c[48]$$

so that $c = 1/48$

Therefore, $f(x) = \frac{1}{48}(x+3),\ 2 \leq x \leq 8$

(b) $$P(3 < X < 5) = \int_3^5 \frac{1}{48}(x+3)dx = \frac{1}{48}\left[\frac{x^2}{2} + 3x\right]_3^5$$

$$= \frac{1}{48}\left[\frac{(5)^2}{2} + 3(5) - \frac{(3)^2}{2} - 3(3)\right] = \frac{1}{48}\left[\frac{25}{2} + 15 - \frac{9}{2} - 9\right]$$

$$= \frac{1}{48}[14] = \frac{7}{24}$$

(c) $P(X \geq 4) = \int_4^8 \frac{1}{48}(x+3)\,dx = \frac{1}{48}\left[\frac{x^2}{2} + 3x\right]_4^8$

$$= \frac{1}{48}\left[\frac{(8)^2}{2} + 3(8) - \frac{(4)^2}{2} - 3(4)\right] = \frac{1}{48}\,[32 + 24 - 8 - 12]$$

$$= \frac{1}{48}\,[36] = \frac{3}{4}$$

Example 8.11.

(a) A continuous random variable X has a density function $f(x) = 2x$ when $0 \leq x \leq 1$ and zero otherwise. Find (i) $P\left(X < \frac{1}{2}\right)$ (ii) $P\left(\frac{1}{4} < X < \frac{1}{2}\right)$

(b) If f(x) has probability density kx^2, $0 < x < 1$, determine k and find the probability that $1/3 < X < 1/2$.

Solution:

(a) $f(x) = 2x,\ 0 \leq x \leq 1,$

$\quad\ \ = 0,$ otherwise.

(i) $P\left(X < \frac{1}{2}\right) = \int_0^{1/2} f(x)\,dx = \int_0^{1/2} 2x\,dx = 2\left[\frac{x^2}{2}\right]_0^{1/2} = \left[\left(\frac{1}{2}\right)^2 - 0\right] = \frac{1}{4}$

(ii) $P\left(\frac{1}{4} < X < \frac{1}{2}\right) = \int_{1/4}^{1/2} f(x)\,dx = \int_{1/4}^{1/2} 2x\,dx = 2\left[\frac{x^2}{2}\right]_{1/4}^{1/2}$

$$= \left[\left(\frac{1}{2}\right)^2 - \left(\frac{1}{4}\right)^2\right] = \left[\frac{1}{4} - \frac{1}{16}\right] = \frac{3}{16}$$

(b) f(x) will be a probability density function, if $\int_{-\infty}^{\infty} f(x)\,dx = 1$, i.e.

$$1 = \int_0^1 f(x)\,dx = \int_0^1 kx^2\,dx = k\left[\frac{x^3}{3}\right]_0^1 = k\left[\frac{1}{3} - 0\right] = \frac{k}{3}, \text{ so that}$$

$k = 3$. Hence the probability density function $f(x) = 3x^2,\ 0 < x < 1$

Now

$P\left(\frac{1}{3} < X < \frac{1}{2}\right) = \int_{1/3}^{1/2} f(x)\,dx = \int_{1/3}^{1/2} 3x^2\,dx = 3\left[\frac{x^3}{3}\right]_{1/3}^{1/2}$

$$= \left[\left(\frac{1}{2}\right)^3 - \left(\frac{1}{3}\right)^3\right] = \left[\frac{1}{8} - \frac{1}{27}\right] = \frac{19}{216}$$

8.13. MATHEMATICAL EXPECTATION

Suppose a random variable X takes the n values as $x_1, x_2, x_3, \ldots, x_n$ with corresponding probabilities $f(x_1), f(x_2), f(x_3), \ldots, f(x_n)$. The mathematical expectation or expected value of X denoted by E(X) is defined as

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + x_3 f(x_3) + \ldots + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

Similarly $$E(X^2) = x_1^2 f(x_1) + x_2^2 f(x_2) + x_3^2 f(x_3) + \ldots + x_n^2 f(x_n) = \Sigma x_i^2 f(x_i)$$

It is also called the mean of the discrete random variable X. If the random experiment is repeated a large number of times, most of the random experiments would generate the result equal to the expected value of X.

8.14. FUNCTION OF A RANDOM VARIABLE

When $Y = aX + b$, it is called linear transformation. The random variable Y is function of the random variable X. Expected value of Y can be written in terms of the expected value of X.

Thus $$E(Y) = E[aX + b] = aE(X) + b$$

It is important to note that the probability function of random variable Y will be the same as the probability function of random variable X. Thus $f(x) = f(y)$.

8.15. LAWS OF EXPECTATION

(i) E(Constant) = Constant

If a constant 'c' is written on all the faces of the die, we shall always get c whenever die is rolled. Thus $E(c) = c$

(ii) $E(aX) = aE(X) \quad a \neq 0$

(iii) $E(X + a) = E(X) + a$

(iv) $E(aX + b) = aE(X) + b \quad a \neq 0$

(v) $E(X + Y) = E(X) + E(Y)$

(vi) $E(X - Y) = E(X) - E(Y)$

(vii) $E(XY) = E(X)E(Y)$ if X and Y are independent.

(viii) $E[X - E(X)] = E(X) - E(X) = 0$

Example 8.12.

Let X have the following probability distribution:

x	1	2	3	4	5	6
p(x)	0.05	0.40	0.10	0.25	0.05	0.15

(i) Find $E(X)$, $E(X^2)$ and $E(X + 4)$

(ii) Verify that $E(2X + 3) = 2E(X) + 3$

Solution:

The necessary calculations are given below:

x	p(x)	x p(x).	x^2 p(x)	x + 4	(x + 4) p(x)	2x + 3	(2x+3) p(x)
1	0.05	0.05	0.05	5	0.25	5	0.25
2	0.40	0.80	1.60	6	2.40	7	2.80
3	0.10	0.30	0.90	7	0.70	9	0.90
4	0.25	1.00	4.00	8	2.00	11	2.75
5	0.05	0.25	1.25	9	0.45	13	0.65
6	0.15	0.90	5.40	10	1.50	15	2.25
Total	1	3.3	13.2		7.3		9.6

(i) $E(X) = \Sigma x\, p(x) = 3.3$

$E(X^2) = \Sigma x^2\, p(x) = 13.2$

$E(X+4) = \Sigma (x+4)\, p(x) = 7.3$

(ii) $E(2X+3) = \Sigma (2x+3)\, p(x) = 9.6$

$2E(X)+3 = 2(3.3)+3 = 6.6+3 = 9.6$

Hence, $E(2X+3) = 2E(X)+3$

Example 8.13.

Two unbiased dice are thrown. Find the expected value of the sum of numbers of points on them.

Solution:

Clearly X may be at least 2 and at the most 12. If $x_1, x_2, x_3, \ldots, x_{11}$ are the values of X with probabilities $p_1, p_2, p_3, \ldots, p_{11}$ respectively corresponding to X = 2, 3, 4, ..., 12, then the probabilities with variate-values may be tabulated as below:

x	2	3	4	5	6	7	8	9	10	11	12	Total
p(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1
x p(x)	2/36	6/36	12/36	20/36	30/36	42/36	40/36	36/36	30/36	22/36	12/36	252/36

$E(X) = \Sigma x\, p(x) = 252/36 = 7$

Example 8.14.

From an urn containing 3 red and 2 white balls, a man is to draw 2 balls at random without replacement, being promised Rs.20 for each red ball he draws, and Rs.10 for each white one. Find his expectation.

Solution:

	Red balls	White balls	Total balls
urn:	3	2	5

S contains $\binom{5}{2} = 10$ sample points

Let x_1 = two red balls = Rs.40

x_2 = one red and one white ball = Rs.30

x_3 = two white balls = Rs.20

The respective probabilities are:

$$p(x_1) = \text{P(two red balls)} = \frac{\binom{3}{2}\binom{2}{0}}{\binom{5}{2}} = \frac{3}{10}$$

$$p(x_2) = \text{P(one red and one white bal l)} = \frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}} = \frac{6}{10}$$

$$p(x_3) = \text{P(two white balls)} = \frac{\binom{3}{0}\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10}$$

$$E(X) = x_1\,p(x_1) + x_2\,p(x_2) + x_3\,p(x_3) = 40\left(\frac{3}{10}\right) + 30\left(\frac{6}{10}\right) + 20\left(\frac{1}{10}\right)$$

$$= 12 + 18 + 2 = \text{Rs.}32$$

Example 8.15.

A and B throw one die for a prize of Rs.77 which is to be won by the player who first throws 3. If A has first throw what are their respective expectations?

Solution:

Here, $p = \frac{1}{6}$ and $q = 1 - p = \frac{5}{6}$

A has the first throw. Therefore A can win in the first trial. If A does not win on the first trial, the second trial will go to B. If B does not win on the second trial, the third trial will go to A and so on. Thus A can win on 1st, 3rd, 5th ... trials and B can win on 2nd, 4th, 6th ... trials.

Trials:	1	3	5	7	...
Chances of A:	p	$p\,q^2$	$p\,q^4$	$p\,q^6$	...
Chances of A:	$\left(\frac{1}{6}\right)$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^6$	...

A may win in any of the trials given to him.

$$P(A) = \left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 + \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4 + \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^6 + \ldots \infty \text{ terms}$$

This is a geometric progression in which $a = \frac{1}{6}$ and $r = \left(\frac{5}{6}\right)^2 = \frac{25}{36}$

$$P(A) = \text{Sum} = \frac{a}{1-r} = \frac{1/6}{1 - 25/36} = \frac{1/6}{11/36} = \frac{6}{11}$$

Trials:	2	4	6	8	...
Chances of B:	$p\,q$	$p\,q^3$	$p\,q^5$	$p\,q^7$	...
Chances of B:	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^5$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^7$	...

B may win in any of the trials given to him.

$$P(B) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)+\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3+\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^5+\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^7+\ldots \infty \text{ terms}$$

This is also a geometric progression in which

$$a = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \frac{5}{36} \quad \text{and} \quad r = \left(\frac{5}{6}\right)^2 = \frac{25}{36}$$

$$P(B) = \text{Sum} = \frac{a}{1-r} = \frac{5/36}{1-25/36} = \frac{5/36}{11/36} = \frac{5}{11}$$

Probability of B can also be obtained from the equation $P(A) + P(B) = 1$

Thus $P(B) = 1 - P(A) = 1 - 6/11 = 5/11$

So that $E(A) = 77 \times 6/11 =$ Rs.42 and $E(B) = 77 \times 5/11 =$ Rs.35

Example 8.16.

In a summer season, a dealer of desert room coolers can earn Rs.800 per day if the day is hot and can earn Rs.300 per day if it is fair and loses Rs.40 per day if it is cloudy. Find his mathematical expectation if the probability of the day being hot is 0.50 and for being cloudy it is 0.30.

Solution:

Let the random variable X denote the number of rupees the dealer earns. Then the possible values of x are 800, 300 and -40, where -40 corresponds to the fact that dealer loses, and the respective probabilities are 0.50, 0.20 and 0.30. Therefore

$$E(X) = 800\,(0.50) + 300(0.20) - 40(0.30) = 400 + 60 - 12 = \text{Rs.}448$$

8.16. VARIANCE AND STANDARD DEVIATION

By definition

$$\text{Var}(X) = \sigma^2 = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

and $$\text{S.D.}(X) = \sqrt{E(X^2) - [E(X)]^2}$$

Example 8.17.

A random variable X has the probability distribution:

$$f(x) = k\binom{5}{x} \quad x = 0, 1, 2, 3, 4, 5.$$

(a) Determine the value of k.

(b) Determine the expected value of X.

(c) Determine the variance of X.

(d) Show that $\text{Var}(4X + 9) = 16\text{Var}(X)$.

Solution:

(a) Since the sum of probabilities is one.

$$\sum_{x=0}^{5} f(x) = 1 \text{ gives}$$

$$k\left[\binom{5}{0}+\binom{5}{1}+\binom{5}{2}+\binom{5}{3}+\binom{5}{4}+\binom{5}{5}\right] = 1$$

$$k[\,1+5+10+10+5+1\,] = 1 \text{ or } k[\,32\,] = 1, \text{ so that } k = 1/32$$

Thus, $f(x) = \frac{1}{32}\binom{5}{x} \quad x = 0, 1, 2, 3, 4, 5.$

x	f(x)	x f(x)	x^2 f(x)	(4x + 9)	(4x + 9) f(x)	$(4x + 9)^2$ f(x)
0	1/32	0	0	9	9/32	81/32
1	5/32	5/32	5/32	13	65/32	845/32
2	10/32	20/32	40/32	17	170/32	2890/32
3	10/32	30/32	90/32	21	210/32	4410/32
4	5/32	20/32	80/32	25	125/32	3125/32
5	1/32	5/32	25/32	29	29/32	841/32
Total	1	80/32	240/32	--	608/32	12192/32

(b) $E(X) = \Sigma x\, f(x) = \frac{80}{32} = 2.5$

$E(X^2) = \Sigma x^2 f(x) = \frac{240}{32} = 7.5$

(c) $Var(X) = E(X^2) - [E(X)]^2 = 7.5 - (2.5)^2 = 1.25$

$E(4X+9) = \Sigma(4x+9)\, f(x) = \frac{608}{32} = 19$

$E(4X+9)^2 = \Sigma(4x+9)^2 f(x) = \frac{12192}{32} = 381$

(d) $Var(4X+9) = E(4X+9)^2 - [E(4X+9)]^2 = 381 - (19)^2 = 20$

$16\, Var(X) = 16(1.25) = 20$

Hence, $Var(4X+9) = 16\, Var(X)$

SHORT DEFINITIONS

Random Variable

A random variable is a numerical quantity whose value depends on chance.

or

A rule that assigns a numerical value to each outcome of a sample space is called random variable.

Probability Distribution

A probability distribution gives the probability for each value of the random variable.

or

A list containing both the values of the random variable and their corresponding probabilities is called the probability distribution of the random variable.

Discrete Sample Space

A sample space is called a discrete sample space if it contains a finite or countable number of sample points.

Discrete Random Variable

A discrete random variable is one which can assume any set of possible values which can be counted or listed.

or

When the values of the random variable are countable, it is called a discrete random variable.

Discrete Probability Distribution

A probability distribution is called a discrete probability distribution if it refers to a discrete sample space.

Probability Function

A function denoted by f(x) that can provide the probability for each value of the discrete random variable is called a probability function.

Properties of the Discrete Probability Function

The probability function for a discrete random variable must have the following two properties:

(i) The range of f(x) is the set of real numbers from 0 to 1, inclusive. That is $0 \le f(x) \le 1$.

(ii) The sum of the probabilities assigned to the values of a random variable must be equal to one, that is, $\Sigma f(x) = 1$.

Continuous Sample Space

If the sample space of an experiment is the real line or intervals on the real line, we call it continuous.

Continuous Random Variable

A continuous random variable is one which can assume an infinite spectrum of different values across an interval and which cannot be counted or listed.

or

A variable that can assume any possible value between two points is called a continuous random variable.

Probability Density Function

A probability density function is a formula, or equation, used to represent the probability distribution of a continuous random variable.

or

The function that defines the probability distribution of a continuous random variable is called probability density function.

Mathematical Expectation or Expected Value of X

The expected value of a discrete random variable X is equal to the sum of the products of each value of x and the corresponding value p(x) i.e; $E(X) = \Sigma x\, p(x)$

or

The expected value or mean of a random variable X is defined as follows:

(i) If X is a discrete random variable and f(x) is the probability of x, then the expected value of this random variable is $E(X) = \Sigma x\, f(x)$.

(ii) If X is a continuous random variable and f(x) is the value of its probability density at x, then the expected value of this random variable is

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx.$$

MULTIPLE – CHOICE QUESTIONS

1. If in a table all possible values of a random variable are given their corresponding probabilities, then this table is called as:
   (a) probability density function (b) distribution function
   (c) probability distribution (d) continuous distribution
2. A variable that can assume any possible value between two points is called:
   (a) discrete random variable (b) continuous random variable
   (c) discrete sample space (d) random variable
3. A formula or equation used to represent the probability distribution of a continuous random variable is called:
   (a) probability distribution (b) distribution function
   (c) probability density function (d) mathematical expectation
4. If X is a discrete random variable and f(x) is the probability of X, then the expected value of this random variable is equal to:
   (a) $\Sigma f(x)$ (b) $\Sigma[x+f(x)]$
   (c) $\Sigma f(x) + x$ (d) $\Sigma x f(x)$
5. Given $E(X) = 5$ and $E(Y) = -2$, then $E(X - Y)$ is:
   (a) 3 (b) 5
   (c) 7 (d) -2
6. Given $x = 2$ and $f(x) = 0.5$. If $y = 2x - 3$, then $f(y)$ is equal to:
   (a) 1 (b) 0.5
   (c) -2 (d) 0

7. Which of the following is not possible in probability distribution?
   (a) $p(x) \geq 0$ (b) $\Sigma p(x) = 1$
   (c) $\Sigma x p(x) = 2$ (d) $p(x) = -0.5$
8. If **c** is a constant (non-random variable), then E(C) is:
   (a) 0 (b) 1
   (c) cf(c) (d) c
9. A discrete probability distribution may be represented by:
   (a) table (b) graph
   (c) mathematical equation (d) all of the above
10. A probability density function may be represented by:
   (a) table (b) graph
   (c) mathematical equation (d) both (b) and (c)
11. If C is a constant in a continuous probability distribution, then $p(x = C)$ is always equal to:
   (a) zero (b) one
   (c) negative (d) impossible
12. $E[X - E(X)]$ is equal to:
   (a) E(X) (b) Var(X)
   (c) 0 (d) $E(X) - X$
13. $E[X - E(X)]^2$ is:
   (a) E(X) (b) $E(X^2)$
   (c) Var(X) (d) S.D.(X)
14. If the random variable takes negative values, then the negative values will have:
   (a) positive probabilities (b) negative probabilities
   (c) constant probabilities (d) difficult to tell
15. If we have $f(x) = 2x$, $0 \leq x \leq 1$, then f(x) is a:
   (a) probability distribution (b) probability density function
   (c) distribution function (d) continuous random variable
16. Numbers selected by a random process and are equally distributed in a table are called:
   (a) attributes (b) random variables
   (c) random numbers (d) quantitative variables
17. $\int_{-\infty}^{\infty} f(x)\, dx$ is always equal to:
   (a) zero (b) one
   (c) E(X) (d) $f(x) + 1$

18. A listing of all the outcomes of an experiment and the probability associated with each outcome is called:
   (a) probability distribution (b) probability density function
   (c) attributes (d) distribution function

19. A quantity resulting from an experiment that, by chance, can assume different values is called:
   (a) random experiment (b) random sample
   (c) random variable (d) random process

20. Which one is not an example of random experiment?
   (a) A coin is tossed and the outcome is either a head or a tail
   (b) A six-sided die is rolled
   (c) Some number of persons will be admitted to a hospital emergency room during any hour.
   (d) All medical insurance claims received by a company in a given year.

21. A set of numerical values assigned to a sample space is called:
   (a) random sample (b) random variable
   (c) random numbers (d) random experiment

22. A variable which can assume finite or countably infinite number of values, is known as:
   (a) continuous (b) discrete
   (c) qualitative (d) none of them

23. The probability function of a random variable is defined as:

x	−1	−2	0	1	2
f(x)	k	2k	3k	4k	5k

   Then k is equal to:
   (a) zero (b) 1/4
   (c) 1/15 (d) one

24. If $f(x) = 1/10$, $x = 10$, then E(X) is:
   (a) zero (b) 6/8
   (c) 1 (d) −1

25. If $\text{Var}(X) = 5$ and $\text{Var}(Y) = 10$, then $\text{Var}(2X + Y)$ is:
   (a) 15 (b) 20
   (c) 10 (d) 30

26. A discrete probability function f(x) is always:
   (a) non-negative (b) negative
   (c) one (d) zero

27. In a discrete probability distribution the sum of all the probabilities is always equal to:
    (a) zero (b) one
    (c) minimum (d) maximum
28. The suitable graph of probability function of a discrete random variable is:
    (a) curve (b) polygon
    (c) probability histogram (d) historigram
29. The appropriate graph of probability density function is.
    (a) curve (b) histogram
    (c) polygon (d) none of them
30. A variable which can assume all values in the range of a random variable, is called:
    (a) finite (b) infinite
    (c) continuous (d) discrete
31. Total area under the curve of a continuous probability density function is always equal to:
    (a) zero (b) one
    (c) -1 (d) none of them
32. An expected value of a random variable is equal to its:
    (a) variance (b) standard deviation
    (c) mean (d) covariance
33. The probability of a continuous random variable "X" taking any particular value, **k** is always:
    (a) negative (b) zero
    (c) one (d) none of them
34. Area of a trapezoid is equal to:
    (a) $\frac{\text{perpendicular} \times \text{base}}{2}$ (b) length × breadth
    (c) $\frac{\text{sum of parallel sides} \times \text{base}}{2}$ (d) none of them

$\mathrm{Var}(4X + 8)$ is:
    (a) $12\,\mathrm{Var}(X)$ (b) $4\,\mathrm{Var}(X) + 8$
    (c) $16\,\mathrm{Var}(X)$ (d) $16\,\mathrm{Var}(X) + 8$

$\mathrm{Var}(X)$ is equal to:
    (a) $E(X^2)$ (b) $[E(X)]^2$
    (c) $E(X^2) - [E(X)]^2$ (d) $E(X^2)+[E(X)]^2$

The expectation of the sum of two random variables X and Y is equal to:
    (a) $E(X)\,E(Y)$ (b) $E(X) + E(Y)$
    (c) $E(X \pm Y)$ (d) $E(XY)$

38. The expectation of the product of two independent variables X and Y is equal to:
   (a) $E(X)\,E(Y)$ (b) $E(X) \pm E(Y)$
   (c) $E(X+Y)$ (d) none of the above
39. When the random variable X and Y are independent, its co-variance is:
   (a) one (b) negative
   (c) zero (d) positive
40. A discrete probability function f(x) is always non-negative and always lies between:
   (a) 0 and ∞ (b) 0 and 1
   (c) -1 and $+1$ (d) $-\infty$ to $+\infty$
41. The probability density function p(x) cannot exceed:
   (a) zero (b) one
   (c) mean (d) infinity
42. The height of persons in a country is a random variable of the type:
   (a) discrete random variable (b) continuous random variable
   (c) both (a) and (b) (d) neither (a) and (b)
43. A random variable is also called:
   (a) chance variable (b) stochastic variable
   (c) constant (d) both (a) and (b)
44. The distribution function F(x) is equal to:
   (a) $P(X = x)$ (b) $P(X \leq x)$
   (c) $P(X \geq x)$ (d) all of the above

Answers

1.	(c)	2.	(b)	3.	(c)	4.	(d)	5.	(c)	6.	(b)	7.	(d)	8.	(d)
9.	(d)	10.	(d)	11.	(a)	12.	(c)	13.	(c)	14.	(a)	15.	(b)	16.	(c)
17.	(b)	18.	(a)	19.	(c)	20.	(d)	21.	(b)	22.	(b)	23.	(c)	24.	(c)
25.	(d)	26.	(a)	27.	(b)	28.	(c)	29.	(a)	30.	(c)	31.	(b)	32.	(c)
33.	(b)	34.	(c)	35.	(c)	36.	(c)	37.	(b)	38.	(a)	39.	(c)	40.	(b)
41.	(b)	42.	(b)	43.	(d)	44.	(b)								

SHORT QUESTIONS

Q.1 What is a random variable?

Q.2 Describe the two different types of random variable.

Q.3 Define a continuous random variable.

Q.4 What is the sample space of a random experiment?

Q.5 Define a probability distribution for a discrete random variable.

or

What is meant by probability distribution?

Q.6 Explain the probability function.

Q.7 Write short note on discrete random variable.

Q.8 Present the graph of discrete probability function.

Q.9 Define the term discrete probability function.

Q.10 What are the properties associated with a discrete probability distribution?

or

Write down the properties of discrete probability distribution.

Q.11 What is meant by mathematical expectation of a random variable?

Q.12 Write down the laws of mathematical expectation.

or

Write down the properties of expected values.

Q.13 Define a random variable and its expectation.

Q.14 Which formulas should be used to calculate the mean, variance and standard deviation of discrete probability distribution?

Q.15 What is a probability density function?

Q.16 Write down the properties of probability density function.

Q.17 Distinguish between discrete and continuous random variables giving examples.

Q.18 What do you mean by expected value?

Q.19 What does " p.d.f " stand for?

Q.20 What are the expectation and standard deviation of a constant?

Ans. $E(C) = C$ and $S.D.(C) = 0$

Q.21 Explain the probability mass function.

Q.22 Write down the formula for the probability distribution of the sum of spots when a pair of dice is thrown.

Ans. $p(x) = \dfrac{6 - |7 - x|}{36}$ for $x = 2, 3, 4, \ldots, 12$.

Q.23 Explain the application of random numbers.

Q.24 Which methods are used in drawing random numbers?

Q.25 Given $f(x) = \dfrac{x}{10}$ $x = 1, 2, 3, 4$. Show that $f(x)$ is a probability function.

Ans. $\Sigma f(x) = 1$

Q.26 Given $x = 0, 1, 2$ and $p(x) = 5/8, 4/8, 1/8$. Is this probability distribution?

Ans. No because $\sum p(x) = 1.25$

Q.27 Given $f(x) = \dfrac{k}{x}$ $x = 1, 2, 3$. Find k.

Ans. 6/11

Q.28 Given $p(x) = k\dbinom{4}{x}$ and $x = 0, 1, 2, 3, 4$. Find the value of k.

Ans. 1/16

Q.29 Given $x = -10, -20, 30$ and $p(x) = 2k, 3k, 5k$. Find $P(X < 30)$.

Ans. 0.5

Q.30 Given $x = -3, -2, 2, 3$ and $p(x) = c, 4c, 3c, 2c$. Find $P(X > -3)$.

Ans. 0.9

Q.31 Given $x = -3, -1, 0, 1, 2, 5$ and $p(x) = 0.1, c, 0.2, 2c, 0.3, 3c$. Find $P(-1 \le X \le +1)$.

Ans. 0.4

Q.32 What would be the probability distribution for the outcomes associated with the throw of a fair six-sided die?

Ans.

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

Q.33 Given $x = 1, 2, 3, 4, 5$ and $p(x) = 1/10, 3/10, p, 2/10, 1/10$. Find the value of p.

Ans. 3/10

Q.34 Three coins are tossed if x = number of tails, then make the probability distribution of x.

Ans.

x	0	1	2	3	Total
p(x)	1/8	3/8	3/8	1/8	1

Q.35 Two coins are tossed, if x = number of heads, then make the probability distribution of x.

Ans.

x	0	1	2	Total
p(x)	1/4	2/4	1/4	1

Q.36 Given Head : Tail = 4 : 2. Find P(Head) and P(Tail).

Ans. P(Head) = 2/3 and P(Tail) = 1/3

Q.37 A pair of dice is thrown, find the expected value of the number of sixes after making its probability distribution.

Ans. 1/3

Q.38 Given $x = -1, 0, 1$ and $p(x) = \frac{2x+3}{4}$. Whether this is a probability distribution or not?

Ans. No because $\sum p(x) = 2.25$

Q.39 Given $x = 2, 4, 6$ and $p(x) = 0.3, Y, 0.2$. Find the value of Y.

Ans. 0.5

Q.40 Given $x = -1, 0, 1, 2$ and $p(x) = k\,|x - 2|$. Find $P(X = 0)$.

Ans. 1/3

Q.41 Given $x = 0, 1, 2$, and $p(x) = 5/40, 25/40, 10/40$. Find $E(X)$.

Ans. 1.125

Q.42 Given $f(x) = \frac{k}{x^2}$ $x = 1, 2$. Find $E(X)$.

Ans. 1.2

Q.43 Let the random variable X have the probability function:

$f(x) = \frac{(|x| + 1)^2}{9}$ for $x = -1, 0, 1$. Compute $E(X)$.

Ans. zero

Q.44 What is the mathematical expectation if we stand to win Rs. 10 if head falls on a balanced coin?

Ans. Rs.5

Q.45 What is the mathematical expectation to win Rs. 50 if a head appears and lose Rs. 50 if a tail appears?

Ans. zero

Q.46 Given $x = 1, 2, 3$ and $p(x) = a, b, a$ where **a** and **b** are constants. Find $E(X)$

Ans. 2

Q.47 Given $x = -1, -2, 0, 1, 2$ and $p(x) = k, 2k, 3k, 4k, 5k$. Find $E(X)$.

Ans. 0.6

Q.48 If a pair of fair dice is rolled and X denotes the sum of the numbers on them, then find the expected value of X.

Ans. 7

Q.49 A die is rolled if x = face value, then find $E(X)$.

Ans. 3.5

Q.50 Given the discrete probability distribution $p(x) = (x^2 + 1)/5$ for $x = -1, 0, 1$. Find $E(X)$.

Ans. 0

Q.51 Given $p(x) = 1/n$ and $x = 1, 2, 3, \ldots, n$. Find $E(X)$.

Ans. $(n + 1)/2$

Q.52 A and B enter into a bet according to which A will get Rs.1000 if it rains on that day and will lose Rs.500 if it does not rain. The probability of raining on a day is 0.7. What is the mathematical expectation of A?

Ans. Rs. 550

Q.53 Given $p(x) = c|x - 2|$ and $x = -1, 0, 1, 3$. Find $E(X)$.

Ans. 1/7

Q.54 Given $x = 0, 2, 3$ and $p(x) = \frac{|1 - x|}{4}$. Find $E(X)$.

Ans. 2

Q.55 Given the discrete probability distribution $p(x) = \binom{3}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{3-x}$ $x = 0,1,2,3$.

Find $E(X^2)$.

Ans. 3

Q.56 Given $x = 1, 2, 3$ and $p(x) = 5/8, 2/8, 1/8$. Find $E(X + 5)$.

Ans. 6.5

Q.57 Given $x = 2, 4, 6$ and $p(x) = 2/6, 2/6, 2/6$. Find $E(2X)$.

Ans. 8

Q.58 Given x = 1, 2, 3 and p(x) = 3/10, 6/10, 1/10. Find $E(2X^2)$.

Ans. 7.2

Q.59 Given x = 2, 3, 4, 5 and p(x) = 3/10, 2/10, 3/10, 2/10. Find $E[X - E(X)]$.

Ans. 0

Q.60 Given x = 1, 2, 3, 4 and p(x) = 1/8, 4/8, 1/8, 2/8. Find $E(X^2 + 2)$.

Ans. 9.25

Q.61 Find the mean and standard deviation for the following discrete distribution:

$$f(x) = 1 \text{ and } x = 5$$

Ans. 5 ; 0

Q.62 Given $E(X) = 0$ and $E(X)^2 = 8/9$, find $E(3X^2 - 2X + 4)$.

Ans. 6.67

Q.63 What is the expectation of a person who is to get Rs.240 if he obtains 3 heads in a single toss of three coins.

Ans. 30

Q.64 If it rains a raincoat dealer can earn Rs.500 per day. If it is fair he can lose Rs. 100 per day. What is his standard deviation if the probability of rain is 0.4?

Ans. Rs. 272.76

Q.65 Given x = 0, 1, 2 and p(x) = 9/16, 6/16, 1/16. Find Var(X).

Ans. 0.375

Q.66 Given E(X) = 0.5, Var(X) = 1.44 and Y = 2X + 2. Find E(Y) and S.D (Y).

Ans. (3, 2.4)

Q.67 Given E(X) = 120 and Var (X) = 100. If $Y = \frac{X-120}{10}$, then find mean and variance of Y.

Ans. (0, 1)

Q.68. Two coins are tossed. If **x** denotes the number of heads. Find E(X).

Ans. 1

Q.69 Given $E(5X + 10) = 18$ and $E(5X + 10)^2 = 524$. Find Var(X).

Ans. 8

Q.70 Given $E(X - 2)^2 = 25$ and $Var(X - 2) = 9$. Find E(X)

Ans. 6

Q.71 Given Var(X) = 2 and Var(Y) = 4. If X and Y are independent random variables then show that $Var\left(\frac{X}{2} - \frac{Y}{4}\right) = 0.75$.

Q.72 Given E(X) = 200 and S.D.(X) = 5. Find $E(X^2)$.

Ans. 40025

Q.73 Given E(X) = 5 and Var(X) = 1. Find E(2 – 3X) and Var(2 – 3X).

Ans. – 13 and 9

Q.74 Given $E(2X) = 10$ and $E(2X)^2 = 500$, determine S.D(X).

Ans. 10

Q.75 Given a random variable X with $E(X) = 2.5$ and $Var(X) = 1.25$. Find $E(X^2)$.

Ans. 7.5

Q.76 Given $E(X^2) = 400$ and $S.D.(X) = 12$. Find $E(X)$.

Ans. 16

Q.77 Given $E(X) = 0.55$, $Var(X) = 1.35$ and $Y = 2X + 1$. Find $E(Y)$ and $Var(Y)$.

Ans. 2.1 and 5.4

Q.78 Given $E(X + 4) = 10$ and $E(X + 4)^2 = 116$, determine $Var(X + 4)$.

Ans. 16

Q.79 Given $E(X + 4) = 10$ and $E(X + 4)^2 = 116$, determine $E(X)$ and $Var(X)$.

Ans. 6;16

Q.80 Given $E(X) = 200$, C.V. $(X) = 7\%$. Find $Var(X)$.

Ans. 196

Q.81 A continuous random variable X that can assume values between $x = 0$ and $x = 2$ has a density function given by $f(x) = x/2$. Show that the area under the curve is equal to 1.

Ans. Total Area = 1

Q.82 A continuous random variable X has a density function $f(x) = c(4 - x)$ for $x = 1$ to $x = 3$, zero otherwise. Find **c**.

Ans. 1/4

Q.83 A continuous random variable X that can assume values between $x = 1$ and $x = 3$ has a density function given by $f(x) = 1/2$. Show that the area under the curve is equal to one.

Ans. Total Area = 1

Q.84 A continuous random variable X has probability density function $f(x) = cx$, for $0 < x < 1$ and $f(x) = 0$, elsewhere. Find the value of c.

Ans. 2

Q.85 A continuous random variable X has probability density function $f(x) = 0.1$, for $x = 0$ to 10 and $f(x) = 0$, elsewhere. Find $P(X = 5)$.

Ans. 0

Q.86 A continuous random variable X has a density function $f(x) = \frac{cx}{4}$ for $1 \le x \le 4$. Find the value of **c**.

Ans. 8/15

EXERCISES

Q.1 Three balls are drawn from a bag containing 5 white and 3 black balls. If 'X' denotes the number of white balls drawn from the bag, then find the probability distribution of X.

Ans.

x:	0	1	2	3
p(x):	1/56	15/56	30/56	10/56

Q.2 There are seven candidates for three positions of typist. Four of the candidates know urdu typing, while the other three do not know it. If the three candidates are selected at random, find the probability distribution of the number of persons knowing urdu typing among those selected.

Ans.

x:	0	1	2	3
p(x):	1/35	12/35	18/35	4/35

Q.3 A coin is tossed 4 times. If X denotes the number of tails, what is the probability distribution of X? Draw a probability histogram.

Ans.

x:	0	1	2	3	4
p(x):	1/16	4/16	6/16	4/16	1/16

Q.4 A bag contains 4 red and 6 black balls. A sample of 4 balls is selected from the bag without replacement. Let X be the number of red balls. Find the probability distribution of X.

Ans.

x:	0	1	2	3	4
p(x):	15/210	80/210	90/210	24/210	1/210

Q.5 A random variable X has the probability distribution:

x	0	1	2	3	4	5	6	7	8
p(x)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine the value of **a**. (ii) Find $P(X < 2)$, $P(X \geq 6)$, $P(3 < X < 5)$.

Ans. (i) 1/81 (ii) 4/81, 5/9, 1/9

Q.6 A random variable X has the following probability distribution:

x	0	1	2	3
p(x)	0.1	0.2	0.3	0.4

Find the probability distribution of Y where $Y = X^2 + 1$.

Ans.

y:	1	2	5	10
p(y):	0.1	0.2	0.3	0.4

Q.7 Find the probability distribution of the number of boys in families with three children, assuming equal probabilities for boys and girls.

Ans.

x:	0	1	2	3
p(x):	1/8	3/8	3/8	1/8

Q.8 From a lot containing 12 items, 4 of which are defective, 5 are chosen at random. If X be the number of defectives found in the sample, write down

(i) the probability distribution of X (ii) $P(X \le 1)$

(iii) $P(1 < X < 3)$ (iv) Verify $\sum_{x=0}^{4} p(x) = 1$.

Ans. (i)

x:	0	1	2	3	4
p(x):	7/99	35/99	42/99	14/99	1/99

(ii) 14/33 (iii) 14/33

Q.9 A continuous random variable X has a density function $f(x) = \frac{x+1}{8}$ for $x = 2$ to $x = 4$. Find (i) $P(X < 3.5)$ (ii) $P(2.4 \le X \le 3.5)$ (iii) $P(X = 1.5)$.

Ans. (i) 0.703 (ii) 0.543 (iii) 0

Q.10 A continuous random variable X that can assume values between X = 2 and X = 5 has a density function given by $f(x) = \frac{2}{27}(x+1)$. Find

(i) $P(X < 4)$ (ii) $P(3 < X < 4)$.

Ans. (i) 16/27 (ii) 1/3

Q.11 A continuous random variable X which can assume values between X = 2 and 8 inclusive has a density function given by a $(x + 3)$ where 'a' is a constant. Find

(i) a (ii) $P(3 < X < 6)$ (iii) $P(X \le 6)$ (iv) $P(X \ge 4)$.

Ans. (i) 1/48 (ii) 15/32 (iii) 7/12 (iv) 3/4

Q.12 A continuous random variable X having values only between 0 and 4 has a density function given by.

$f(x) = \frac{1}{2} - ax$, where 'a' is a constant Find: (i) a (ii) $P(1 < X < 2)$

Ans. (i) 1/8 (ii) 5/16

Q.13 A continuous random variable X has probability density function

$f(x) = cx$ for $0 < X < 2$. Determine (i) c (ii) probability that $1 < X < 1.5$

(iii) probability that $X < 1.5$.

Ans. (i) 1/2 (ii) 0.3125 (iii) 0.5625

Q.14 If a function $f(x) = \frac{1}{18}(3 + 2x)$, $2 \le X \le 4$. Show that it is a density function and find the probability that $2 \le X \le 3$.

Ans. 4/9

Q.15 Given the following discrete probability distribution:

x	0	1	2	3	4	5
p(x)	6/36	10/36	8/36	6/36	4/36	2/36

Compute its mean, variance, standard deviation and coefficient of variation.

Ans. 1.944, 2.054, 1.433, 73.71 %

Q.16 Let X be a random variable with probability distribution as follows:

x	1	2	3	4	5
f(x)	0.125	0.45	0.25	0.05	0.125

Show that $E(5X + 8) = 5E(X) + 8$.

Ans. 21

Q.17 Given a random variable X with $E(X) = 5$ and another random variable Y with $E(Y) = 10$. Find $E(X + 2Y)$ and $E(12 - 2X)$.

Ans. 25, 2

Q.18 A random variable X has the following probability distribution:

Value of X	Probability Distribution of X, $P(X = x)$
1	$P(X = 1) = 6/9$
2	$P(X = 2) = 2/9$
3	$P(X = 3) = 1/9$

(i) Determine the expected value of X (ii) Determine the variance of X

(iii) Determine the standard deviation of X.

Ans. (i) 1.444 (ii) 0.471 (iii) 0.686

Q.19 Let X be a random variable with probability distribution:

x	– 1	0	1	2	3
p(x)	0.125	0.500	0.200	0.050	0.125

Find the probability distribution of the random variable $Y = 2X + 1$. Using the probability distribution of Y, determine $E(Y)$ and $Var(Y)$.

Ans.

y:	– 1	1	3	5	7	$E(Y) = 2.1$
p(y):	0.125	0.500	0.200	0.050	0.125	$Var(Y) = 5.39$

Q.20 If $f(x) = \dfrac{6 - |7 - x|}{36}$ for $x = 2, 3, 4, \ldots, 12$, then find the mean and variance of the random variable X.

Ans. 7, 5.83

Q.21 Find (i) $E(X)$ (ii) $E(X^2)$ (iii) $E[X - E(X)]^2$ for the following probability distribution.

x	– 10	– 20	30
p(x)	1/5	3/10	1/2

Ans. (i) 7 (ii) 590 (iii) 541

Q.22 A random variable X takes the values $-3, -2, 2, 3$ and 4 with probabilities p(x) equal to 1/5, 1/10, 1/10, 1/5 and 2/5 respectively. Compute the variance of $(5X + 10)$.

Ans. 206

Q.23 Let X have the following probability distribution:

x_i	1	2	3	4	5
$p(x_i)$	0.2	0.3	0.2	0.2	0.1

Find $E(3X-1)$, $E(X^2)$ and $E(X^2+2)$.

Ans. 7.1, 8.9, 10.9

Q.24 If $E(X) = 5$, $E(X^2) = 36$, find the mean and variance of $2X - 5$.

Ans. 5, 44

Q.25 What is the expectation of a person who is to get Rs.80 if he obtains 3 heads in a single toss of 3 coins.

Ans. Rs.10

Q.26 A committee of size 5 is to be selected at random from 3 women and 5 men. Find the expected number of women on the committee.

Ans. 2 approximately

Q.27 A and B throw a die for a prize of Rs.55, which is to be won by the player who first throws 6. If A has the first throw, what are their respective expectations?

Ans. Rs.30, Rs.25

Q.28 Three balls are drawn from a bag containing 5 white and 3 black balls. If X denotes the number of white balls drawn from the bag, then find the probability distribution of X. Also find its mean and variance.

Ans.

x:	0	1	2	3	$E(X) = 1.875$
p(x):	1/56	15/56	30/56	10/56	$Var(X) = 0.502$

Q.29 If a pair of fair dice is rolled and X denotes the sum of the numbers on them, then find the expectation of X. If only a single die is rolled what is the expectation of a number on the die.

Ans. 7, 3.5

Q.30 A man draws 2 balls from a bag containing 3 white and 5 black balls. If he is to receive Rs.70 for every white ball which he draws and Rs.7 for every black ball, find his expectation.

Ans. Rs.61.25

Q.31 A person plays a game of tossing a die under the condition that he would get twice as many rupees as the number of points on the upper face. Find his expectation.

Ans. 7

Q.32 In a summar season, a dealer of desert room coolers can earn Rs.800 per day if the day is hot and can earn Rs.200 per day if it is fair and loses Rs.50 per day if it is cloudy. Find his mathematical expectation if the probability of the day being hot is 0.40 and for being cloudy it is 0.35.

Ans. Rs.352.5

Q.33 A coin is biased such that a head is thrice as likely to occur as a tail. Find the probability distribution of heads and also find the mean and variance of distribution when it is tossed 4 times.

Ans.

x:	0	1	2	3	4	$E(X) = 3$
p(x):	1/256	12/256	54/256	108/256	81/256	$Var(X) = 0.75$

Q.34 Approximately 10 % of the glass bottles coming from a production line have serious defects. If two bottles are selected at random, find the expected number of bottles that have serious defects.

Ans. $E(X) = 0.2$

Q.35 If it rains a raincoat dealer can earn Rs.500 per day. If it is fair he can lose Rs.100 per day. What is his expectation if the probability of rain is 0.4?

Ans. Rs.140

Q.36 A and B enter into a bet according to which A will get Rs.200 if it rains on that day and will lose Rs.100 if it does not rain. The probability of raining on a day is 0.7. What is the mathematical expectation of A?

Ans. Rs.110

Chapter

9

BINOMIAL AND HYPERGEOMETRIC DISTRIBUTIONS

9.1 INTRODUCTION

In random experiments, we can divide a sample space into two parts. Fruit in a basket can be divided into two parts sweet and sour, in a throw of die the six faces can be divided into even and odd faces or prime and non-prime faces. When two dice are thrown, the total on the dice may be divided into even or odd. These experiments have only two possible outcomes. A very simple example of this type of experiment is throw of a coin which shows a head or a tail. The probability of each outcome is not disturbed by repeating such experiment. In tossing a coin large number of times, the probability of head remains constant and every time the result of coin is independent of previous throws. We apply the binomial distribution in all these situations in which sample space is divided into two parts according to some definition of partition. Independent trials with only two possible outcomes and constant probabilities are called binomial or, sometimes, Bernoulli Trials, after J. Bernoulli (1654 – 1705), who was a pioneer in the field of probability.

9.2 BERNOULLI TRIALS

A student selected from a class may be intelligent or non-intelligent, a bulb selected from a big lot may be good or defective, a person may be smoker or non-smoker. These trials have only two possible outcomes. One of these outcomes is called success and the other is called failure. The word success is not always used for outcomes like "intelligent", "good" or "useful". This word is used for the outcome in which we are interested. It may be used for a non-intelligent student or for a defective bulb. The probability of success is usually denoted by p and the probability of failure is denoted by q ($q = 1 - p$). The term Bernoulli trials is used for these trials provided the probability of success p does not change from trial to trial.

9.3 BINOMIAL EXPERIMENT

An experiment with n independent trials in which the outcomes can always be classified as either a success or a failure and the probability of success remains constant from trial to trial is called a binomial experiment. A binomial experiment has the following properties.

(i) Each trial results in two outcomes which can be classified into success and failure.

(ii) The probability of success remains constant from trial to trial.

(iii) The successive trials are independent.

(iv) The experiment is repeated a fixed number of times say, n.

9.4 BINOMIAL RANDOM VARIABLE

Let X be the total number of successes in n independent Bernoulli trials, with p as the probability of success in a single trial, then X is called the binomial random variable with parameters n and p.

9.5 THE BINOMIAL DISTRIBUTION

The binomial distribution was discovered early in the eighteenth century by the Swiss mathematician, J. Bernoulli (1654 – 1705). It was published after his death in 1713. If p is the probability that an event will happen in any single trial and $q = 1 - p$ is the probability that it will fail to happen in any single trial, then the probability that the event will happen exactly X times in n trials (i.e. x successes and $n - x$ failures will occur) is given by

$$P(X = x) = f(x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x}$$

where $x = 0, 1, 2, 3, \ldots, n$ and $n! = n(n-1)(n-2) \ldots 3.2.1.$ $0! = 1$

The probability distribution of the number of successes so obtained is called the binomial probability distribution for the obvious reason that the probabilities of 0, 1, 2, 3, ..., x, ..., n successes are the respective terms in the binomial expansion $(q + p)^n$. The binomial distribution contains two independent constants, n and p. If n and p are known then we can determine all measures like mean, variance, coefficient of skewness etc. of the distribution. They are called parameters of the binomial distribution. If $p = q = 1/2$, the binomial distribution is a symmetrical distribution and when $p \neq q$, it is a skewed distribution.

9.6 GRAPH OF THE BINOMIAL DISTRIBUTION

The probability function will be symmetrical when $p = 1/2$ (Figure 9.1.). If $p > 1/2$ (Figure 9.2), the probability function will be negatively skewed. If $p < 1/2$ (Figure 9.3), it will be positively skewed. The greater the difference between p and $1 - p$, the greater the skewness of the probability function. However, as n increases, the probability function approaches symmetry regardless of the difference between p and $1 - p$. The degree of skewness can be measured with the help of a formula called

coefficient of skewness = $\dfrac{q-p}{\sqrt{npq}}$.

Figure 9.1. Binomial distribution for $n = 4,\ p = 1/2$.

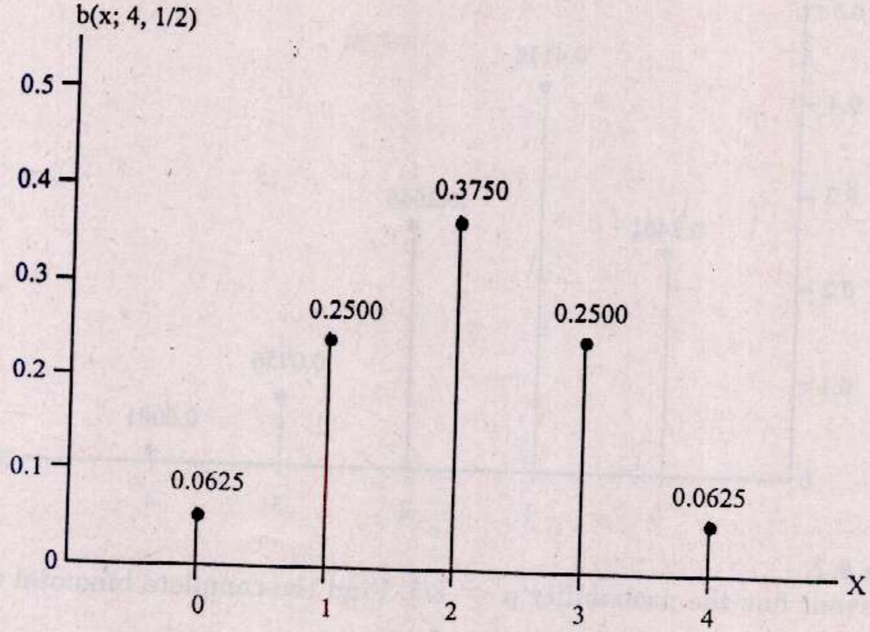


Figure 9.2. Binomial distribution for $n = 4,\ p = 7/10$.

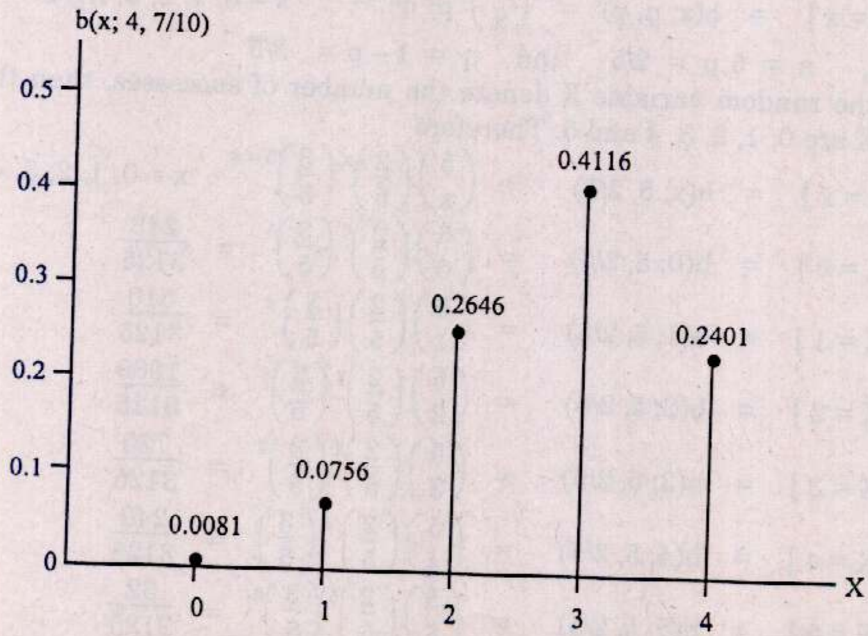
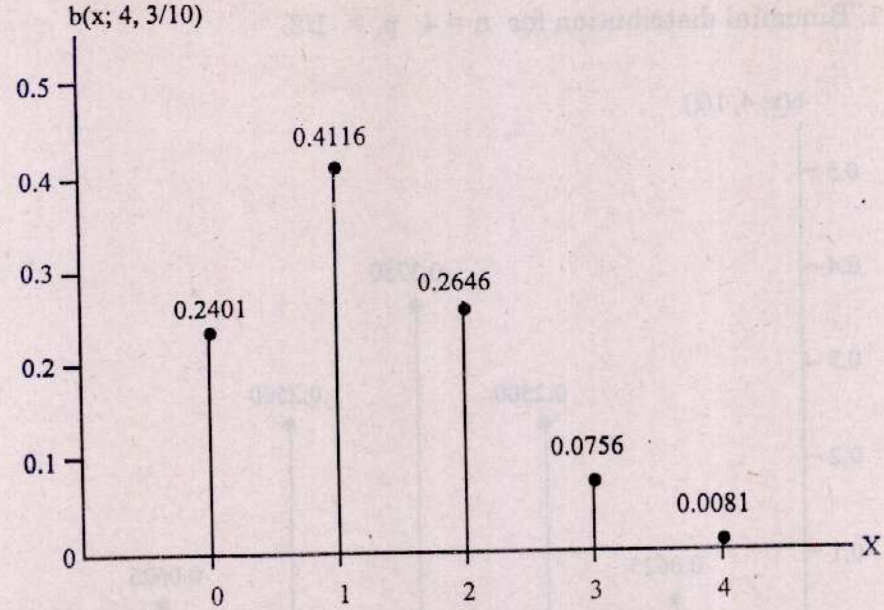


Figure 9.3. Binomial distribution for n = 4, p = 3/10.



Example 9.1.

An event has the probability $p = 2/5$. Find the complete binomial distribution for $n = 5$.

Solution:

The probability of x successes in a series of n trials is given by

$$P[X=x] = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 5, p = 2/5$ and $q = 1 - p = 3/5$

Let the random variable X denote the number of successes, then the possible values of X are 0, 1, 2, 3, 4 and 5. Therefore

$$P[X=x] = b(x; 5, 2/5) = \binom{5}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^{5-x} \qquad x = 0, 1, 2, 3, 4, 5.$$

$$P[X=0] = b(0; 5, 2/5) = \binom{5}{0}\left(\frac{2}{5}\right)^0\left(\frac{3}{5}\right)^5 = \frac{243}{3125}$$

$$P[X=1] = b(1; 5, 2/5) = \binom{5}{1}\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)^4 = \frac{810}{3125}$$

$$P[X=2] = b(2; 5, 2/5) = \binom{5}{2}\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right)^3 = \frac{1080}{3125}$$

$$P[X=3] = b(3; 5, 2/5) = \binom{5}{3}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^2 = \frac{720}{3125}$$

$$P[X=4] = b(4; 5, 2/5) = \binom{5}{4}\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right) = \frac{240}{3125}$$

$$P[X=5] = b(5; 5, 2/5) = \binom{5}{5}\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^0 = \frac{32}{3125}$$

Thus the complete binomial distribution with p = 2/5 and n = 5 in tabular form is given as follows:

x	0	1	2	3	4	5	Total
$P[X=x]$	$\frac{243}{3125}$	$\frac{810}{3125}$	$\frac{1080}{3125}$	$\frac{720}{3125}$	$\frac{240}{3125}$	$\frac{32}{3125}$	$\Sigma P[X=x]=1$

Example 9.2.

The experience of a house-agent indicates that he can provide suitable accommodation for 75 percent of the clients who come to him. If on a particular occasion, 6 clients approach him independently, calculate the probability that:

(i) less than 4 clients will get satisfactory accommodation
(ii) exactly 4 clients will get satisfactory accommodation
(iii) at least 4 clients will get satisfactory accommodation
(iv) at most 4 clients will get satisfactory accommodation.

Solution:

The probability of x successes in a series of n trials is given by

$$P[X=x] = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 6, p = 75/100 = 3/4$, and $q = 1 - p = 1/4$

Let the random variable X denote the number of clients who will get satisfactory accommodation. Then the possible values of X are 0, 1, 2, 3, 4, 5, 6. Therefore

$$P[X=x] = \binom{6}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{6-x} \qquad x = 0, 1, 2, 3, 4, 5, 6.$$

(i) $$P[X<4] = \binom{6}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^6 + \binom{6}{1}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^5 + \binom{6}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^4 + \binom{6}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^3$$

$$= \frac{1}{4096} + \frac{18}{4096} + \frac{135}{4096} + \frac{540}{4096} = \frac{694}{4096} = 0.1694$$

(ii) $$P[X=4] = \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2 = \frac{1215}{4096} = 0.2966$$

(iii) $$P[X\geq 4] = \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2 + \binom{6}{5}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right) + \binom{6}{6}\left(\frac{3}{4}\right)^6\left(\frac{1}{4}\right)^0$$

$$= \frac{1215}{4096} + \frac{1458}{4096} + \frac{729}{4096} = \frac{3402}{4096} = 0.8306$$

(iv) $$P[X\leq 4] = \binom{6}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^6 + \binom{6}{1}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^5 + \binom{6}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^4 + \binom{6}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^3 + \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2$$

$$= \frac{1}{4096} + \frac{18}{4096} + \frac{135}{4096} + \frac{540}{4096} + \frac{1215}{4096} = \frac{1909}{4096} = 0.4661$$

Example 9.3.

Large lots of incoming products at a manufacturing plant are inspected for defectives by means of a sampling scheme. Only 8 items are to be examined and the lot is rejected if 2 or more defectives are observed. If a lot contains 10 % defectives, what is probability that the lot will be:

(i) accepted? (ii) rejected?

Solution:

The probability of x successes in a series of n trials is given by

$$P[X=x] = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 8, p = 10/100 = 1/10$ and $q = 1 - p = 9/10$

Let the random variable X denote the number of defective items, then the possible values of X are 0, 1, 2, 3, ..., 8. Therefore

$$P[X=x] = \binom{8}{x}\left(\frac{1}{10}\right)^x\left(\frac{9}{10}\right)^{8-x} \quad x = 0, 1, 2, 3, \ldots, 8.$$

(i) $P[X<2] = \binom{8}{0}\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^8 + \binom{8}{1}\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^7$

$= 0.4305 + 0.3826 = 0.8131$

(ii) $P[X \geq 2] = 1 - P[X<2] = 1 - 0.8131 = 0.1869.$

Example 9.4

Assuming that each baby has probability 0.4 of being male, find the probability that a family of 4 children will have

(i) exactly one boy (ii) exactly one girl (iii) at least one boy
(iv) at least one girl (v) at most one boy (vi) at most one girl

Solution:

The probability of x successes in a series of n trials is given by

$$P(X=x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \cdots, n.$$

Here, $n = 4, p = 0.4, q = 1 - p = 1 - 0.4 = 0.6$

Let the random variable X denote the number of boys, then the possible values of X are 0, 1, 2, 3, 4. Therefore

$$P(X=x) = \binom{4}{x}(0.4)^x(0.6)^{4-x} \quad x = 0, 1, 2, 3, 4.$$

(i) $P(X=1) = \binom{4}{1}(0.4)^1(0.6)^3 = 0.3456$

(ii) $P(X=3) = \binom{4}{3}(0.4)^3(0.6)^1 = 0.1536$

(iii) $P(X \geq 1) = 1 - P(X=0) = 1 - \binom{4}{0}(0.4)^0(0.6)^4 = 1 - 0.1296 = 0.8704$

(iv) $P(X \le 3) = 1 - P(X = 4) = 1 - \binom{4}{4}(0.4)^4(0.6)^0 = 1 - 0.0256 = 0.9744$

(v) $P(X \le 1) = \binom{4}{0}(0.4)^0(0.6)^4 + \binom{4}{1}(0.4)^1(0.6)^3$

$= 0.1296 + 0.3456 = 0.4752$

(vi) $P(X \ge 3) = \binom{4}{3}(0.4)^3(0.6)^1 + \binom{4}{4}(0.4)^4(0.6)^0$

$= 0.1536 + 0.0256 = 0.1792$

Example 9.5.

Given n = 6, p = 1/3. Find

(a) $P[X = -1]$ (b) $P[X = 2.5]$ (c) $P[X = 2]$ (d) $P[X = 10]$.

Solution:

The probability of x successes in a series of n trials is given by

$$P[X = x] = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 6, p = 1/3, q = 1 - p = 2/3$ and $x = 0, 1, 2, 3, 4, 5, 6$.

$$P[X = x] = \binom{6}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{6-x} \qquad x = 0, 1, 2, 3, \ldots, 6.$$

(a) $P[X = -1] = 0$, because a random variable X in a binomial distribution takes only positive integral values.

(b) $P[X = 2.5] = 0$, because X can take only integer values 0, 1, 2, 3, 4, 5, 6.

(c) $P[X = 2] = \binom{6}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4 = \frac{240}{729}$

(d) $P[X = 10] = 0$, because X can take only values 0, 1, 2, 3, 4, 5, 6.

9.7 MEAN, VARIANCE AND STANDARD DEVIATION OF THE BINOMIAL DISTRIBUTION

The mean, variance and standard deviation of the binomial distribution, means the mean, variance and standard deviation of the values taken by the variable in repeated binomial experiments. Instead of carrying out these experiments in order to calculate the mean, variance and standard deviation, it can be shown mathematically that the following formulae may be applied to any binomial distribution.

$$\text{Mean} = \mu = \Sigma x\, p(x) = np$$

$$\text{Variance} = \sigma^2 = \Sigma x^2 p(x) - [\Sigma x\, p(x)]^2 = npq$$

$$\text{S.D.} = \sigma = \sqrt{\Sigma x^2 p(x) - [\Sigma x\, p(x)]^2} = \sqrt{npq}$$

Example 9.6.

Find the mean, variance and standard deviation of the binomial distribution $(q+p)^2$.

Solution:

Here, $n = 2$, $x = 0, 1, 2$. Therefore

$$P[X=x] = \binom{2}{x} p^x q^{2-x} \qquad x = 0, 1, 2.$$

$$P[X=0] = \binom{2}{0} p^0 q^2 = q^2 \qquad P[X=1] = \binom{2}{1} pq = 2pq$$

$$P[X=2] = \binom{2}{2} p^2 q^0 = p^2$$

Thus, the probability distribution in tabular form for the computation of mean, variance and standard deviation of X is given as follows:

x	$P[X=x]=p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^2	0	0	0
1	$2pq$	$2pq$	1	$2pq$
2	p^2	$2p^2$	4	$4p^2$

$$E(X) = \mu = \Sigma x\,p(x) = 2pq + 2p^2$$

$$= 2p(q+p) = 2p(1) = 2p$$

$$E(X^2) = \Sigma x^2\,p(x) = 2pq + 4p^2$$

$$\text{Var}(X) = \sigma^2 = E(X^2) - [E(X)]^2 = 2pq + 4p^2 - (2p)^2$$

$$= 2pq + 4p^2 - 4p^2 = 2pq$$

$$\text{S.D}(X) = \sigma = \sqrt{2pq}$$

Example 9.7.

Show that the mean = 3p, variance = 3pq and standard deviation = $\sqrt{3pq}$ for a binomial distribution in which n = 3.

Solution:

Here, $n = 3$, $x = 0, 1, 2, 3$. Therefore

$$P[X=x] = \binom{3}{x} p^x q^{3-x} \qquad x = 0, 1, 2, 3.$$

$$P[X=0] = \binom{3}{0} p^0 q^3 = q^3 \qquad P[X=1] = \binom{3}{1} pq^2 = 3pq^2$$

$$P[X=2] = \binom{3}{2} p^2 q = 3p^2 q \qquad P[X=3] = \binom{3}{3} p^3 q^0 = p^3$$

Thus, the probability distribution in tabular form for the computation of mean. variance and standard deviation of X is given as follows:

x	$P[X = x] = p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^3	0	0	0
1	$3pq^2$	$3pq^2$	1	$3pq^2$
2	$3p^2q$	$6p^2q$	4	$12p^2q$
3	p^3	$3p^3$	9	$9p^3$

$$E(X) = \mu = \Sigma x\,p(x) = 3pq^2 + 6p^2q + 3p^3 = 3p(q^2 + 2pq + p^2)$$

$$= 3p(q + p)^2 = 3p(1)^2 = 3p$$

$$E(X^2) = \Sigma x^2\,p(x) = 3pq^2 + 12p^2q + 9p^3 = 3p(q^2 + 4pq + 3p^2)$$

$$= 3p[\,q^2 + 2pq + p^2 + 2pq + 2p^2\,] = 3p[\,(q + p)^2 + 2p(q + p)\,]$$

$$= 3p[\,(1)^2 + 2p(1)\,] = 3p[\,1 + 2p\,] = 3p + 6p^2$$

$$Var(X) = E(X^2) - [\,E(X)\,]^2 = 3p + 6p^2 - (3p)^2 = 3p + 6p^2 - 9p^2$$

$$= 3p - 3p^2 = 3p(1 - p) = 3pq$$

$$S.D(X) = \sigma = \sqrt{3pq}$$

Example 9.8.

Find the mean, variance and standard deviation of the binomial distribution $(q + p)^4$.

Solution:

Here, $n = 4$, $x = 0, 1, 2, 3, 4$. Therefore

$$P[X = x] = \binom{4}{x} p^x q^{4-x} \qquad x = 0, 1, 2, 3, 4.$$

$P[X = 0] = q^4$ $\quad P[X = 1] = 4pq^3$ $\quad P[X = 2] = 6p^2q^2$

$P[X = 3] = 4p^3q$ $\quad P[X = 4] = p^4$

Thus, the probability distribution in tabular form for the computation of mean, variance and standard deviation of X is given as follows:

x	$P[X = x] = p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^4	0	0	0
1	$4pq^3$	$4pq^3$	1	$4pq^3$
2	$6p^2q^2$	$12p^2q^2$	4	$24p^2q^2$
3	$4p^3q$	$12p^3q$	9	$36p^3q$
4	p^4	$4p^4$	16	$16p^4$

$$E(X) = \mu = \Sigma x\, p(x) = 4pq^3 + 12p^2q^2 + 12p^3q + 4p^4$$
$$= 4p(q^3 + 3pq^2 + 3p^2q + p^3) = 4p(q+p)^3 = 4p(1)^3 = 4p$$
$$E(X^2) = \Sigma x^2\, p(x) = 4pq^3 + 24p^2q^2 + 36p^3q + 16p^4$$
$$= 4p[\, q^3 + 6pq^2 + 9p^2q + 4p^3]$$
$$= 4p[\, q^3 + 3pq^2 + 3p^2q + p^3 + 3pq^2 + 6p^2q + 3p^3]$$
$$= 4p[\, (q+p)^3 + 3p(q^2 + 2pq + p^2)]$$
$$= 4p[\, (q+p)^3 + 3p(q+p)^2] = 4p[\, (1)^3 + 3p(1)^2]$$
$$= 4p[\, 1 + 3p] = 4p + 12p^2$$
$$\text{Var}(X) = \sigma^2 = E(X)^2 - [E(X)]^2 = 4p + 12p^2 - (4p)^2$$
$$= 4p + 12p^2 - 16p^2 = 4p - 4p^2 = 4p(1-p) = 4pq$$
$$\text{S.D.}(X) = \sigma = \sqrt{4pq}$$

Example 9.9.

Find the mean, variance and standard deviation of the binomial distribution $(q + p)^n$.

Solution:

The number of successes x will be 0, 1, 2, 3, 4, ..., n. We know that

$$(q+p)^n = \binom{n}{0} p^0q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2q^{n-2} + \binom{n}{3} p^3q^{n-3} + \binom{n}{4} p^4q^{n-4} + \dots + \binom{n}{n} p^nq^0$$

Therefore the probabilities of 0, 1, 2, 3, 4, ..., n successes are respectively,

$$q^n,\ \binom{n}{1} pq^{n-1},\ \binom{n}{2} p^2q^{n-2},\ \binom{n}{3} p^3q^{n-3},\ \binom{n}{4} p^4q^{n-4},\ \dots,\ p^n$$

Thus, the probability distribution in tabular form for the computation of mean, variance and standard deviation of X is given as follows:

x	$P[X = x] = p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^n	0	0	0
1	$\binom{n}{1} pq^{n-1}$	$\binom{n}{1} pq^{n-1}$	1	$\binom{n}{1} pq^{n-1}$
2	$\binom{n}{2} p^2q^{n-2}$	$2\binom{n}{2} p^2q^{n-2}$	4	$4\binom{n}{2} p^2q^{n-2}$
3	$\binom{n}{3} p^3q^{n-3}$	$3\binom{n}{3} p^3q^{n-3}$	9	$9\binom{n}{3} p^3q^{n-3}$
4	$\binom{n}{4} p^4q^{n-4}$	$4\binom{n}{4} p^4q^{n-4}$	16	$16\binom{n}{4} p^4q^{n-4}$
⋮	⋮	⋮	⋮	⋮
n	p^n	np^n	n^2	n^2p^n

$$E(X) = \mu = \Sigma x\, p(x) = \binom{n}{1} pq^{n-1} + 2\binom{n}{2} p^2q^{n-2} + 3\binom{n}{3} p^3q^{n-3} + 4\binom{n}{4} p^4q^{n-4} + \ldots + np^n$$

$$= npq^{n-1} + 2\,\frac{n(n-1)}{2!}\,p^2q^{n-2} + 3\,\frac{n(n-1)(n-2)}{3!}\,p^3q^{n-3} + 4\,\frac{n(n-1)(n-2)(n-3)}{4!}\,p^4q^{n-4} + \ldots + np^n$$

$$= np\left[q^{n-1} + (n-1)pq^{n-2} + \frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \frac{(n-1)(n-2)(n-3)}{3!}p^3q^{n-4} + \ldots + p^{n-1}\right]$$

$$= np\left[q^{n-1} + \binom{n-1}{1}pq^{n-2} + \binom{n-1}{2}p^2q^{n-3} + \binom{n-1}{3}p^3q^{n-4} + \ldots + p^{n-1}\right]$$

$$= np\,[(q+p)^{n-1}] = np\,[(1)^{n-1}] = np$$

$$E(X^2) = \Sigma x^2\, p(x) = \binom{n}{1} pq^{n-1} + 4\binom{n}{2} p^2q^{n-2} + 9\binom{n}{3} p^3q^{n-3} + 16\binom{n}{4} p^4q^{n-4} + \ldots + n^2p^n$$

$$= npq^{n-1} + 4\,\frac{n(n-1)}{2!}\,p^2q^{n-2} + 9\,\frac{n(n-1)(n-2)}{3!}\,p^3q^{n-3} + 16\,\frac{n(n-1)(n-2)(n-3)}{4!}\,p^4q^{n-4} + \ldots + n^2p^n$$

$$= np\left[q^{n-1} + 2(n-1)pq^{n-2} + 3\,\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + 4\,\frac{(n-1)(n-2)(n-3)}{3!}p^3q^{n-4} + \ldots + np^{n-1}\right]$$

$$= np\left[\left\{q^{n-1} + (n-1)pq^{n-2} + \frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \frac{(n-1)(n-2)(n-3)}{3!}p^3q^{n-4} + \ldots + p^{n-1}\right\} + \left\{(n-1)pq^{n-2} + 2\,\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + 3\,\frac{(n-1)(n-2)(n-3)}{3!}p^3q^{n-4} + \ldots + (n-1)p^{n-1}\right\}\right]$$

$$= np\left[(q+p)^{n-1} + (n-1)\,p\left\{(q^{n-2} + (n-2)pq^{n-3} + \frac{(n-2)(n-3)}{2!}p^2q^{n-4} + \ldots + p^{n-2}\right\}\right]$$

$$= np\left[(q+p)^{n-1} + (n-1)\,p\left\{q^{n-2} + \binom{n-2}{1} pq^{n-3} + \binom{n-2}{2} p^2q^{n-4} + \ldots + p^{n-2}\right\}\right]$$

$$= np\,[(q+p)^{n-1} + (n-1)\,p(q+p)^{n-2}] = np\,[(1)^{n-1} + (n-1)\,p(1)^{n-2}]$$

$$= np\,[1 + (n-1)p] = np[\,1 + np - p\,] = np + n^2p^2 - np^2$$

$$\text{Var}(X) = \sigma^2 = E(X^2) - [\,E(X)\,]^2 = np + n^2p^2 - np^2 - (np)^2$$

$$= np + n^2p^2 - np^2 - n^2p^2 = np - np^2 = np(1-p) = npq$$

$$\text{S.D}(X) = \sigma = \sqrt{npq}$$

Alternative method

The binomial random variable X with parameters n and p has the probability distribution

$$P[X = x] = b(x;\, n,\, p) = \binom{n}{x} p^xq^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

$$E(X) = \sum_{x=0}^{n} x\,p(x) = \sum_{x=0}^{n} x\binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x\frac{n!}{x!(n-x)!}\,p^xq^{n-x} = \sum_{x=0}^{n} x\frac{n(n-1)!}{x(x-1)!\,(n-x)!}\,pp^{x-1}q^{n-x}$$

$$= np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!\,(n-x)!}\,p^{x-1}q^{n-x} = np\sum_{x=1}^{n}\binom{n-1}{x-1} p^{x-1}q^{n-x}$$

$$= np\left[q^{n-1} + \binom{n-1}{1} pq^{n-2} + \binom{n-1}{2} p^2q^{n-3} + \ldots + p^{n-1}\right]$$

$$= np[(q+p)^{n-1}] = np\,[\,(1)^{n-1}] = np$$

$$E(X^2) = E[\,X(X-1)\,] + E(X) = E[\,X(X-1)\,] + np$$

Where
$$E[\,X(X-1)\,] = \sum_{x=0}^{n} x(x-1)\,p(x) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\,\frac{n!}{x!(n-x)!}\,p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\,\frac{n(n-1)\,(n-2)!}{x(x-1)\,(x-2)!\,(n-x)!}\,p^2p^{x-2}q^{n-x}$$

$$= n(n-1)\,p^2\sum_{x=2}^{n}\frac{(n-2)!}{(x-2)!\,(n-x)!}\,p^{x-2}\,q^{n-x}$$

$$= n(n-1)\,p^2\sum_{x=2}^{n}\binom{n-2}{x-2} p^{x-2}\,q^{n-x}$$

$$= n(n-1)\,p^2\left[q^{n-2} + \binom{n-2}{1} pq^{n-3} + \binom{n-2}{2} p^2q^{n-4} + \ldots + p^{n-2}\right]$$

$$= n(n-1)p^2\,[(q+p)^{n-2}] = n(n-1)\,p^2[(1)^{n-2}] = n(n-1)\,p^2$$

So, $E(X^2) = n(n-1)p^2 + np = n^2p^2 - np^2 + np$

$$Var(X) = \sigma^2 = E(X^2) - [E(X)]^2 = n^2p^2 - np^2 + np - (np)^2$$
$$= n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p) = npq$$

$$S.D(X) = \sigma = \sqrt{npq}$$

Example 9.10.

If X is a binomial random variable with $E(X) = 1.44$ and $S.D(X) = 0.96$. Find the parameters of the binomial distribution. Also find $P[X = 2]$.

Solution:

Here, $E(X) = np = 1.44$, $S.D(X) = \sqrt{npq} = 0.96$

$Var(X) = npq = (0.96)^2 = 0.9216$

$1.44\,q = 0.9216$ (since $np = 1.44$)

or $q = \dfrac{0.9216}{1.44} = 0.64$ and $p = 1 - q = 0.36$

$np = 1.44$ or $n(0.36) = 1.44$ (since $p = 0.36$)

$n = \dfrac{1.44}{0.36} = 4$

$$P[X = 2] = \binom{4}{2}(0.36)^2(0.64)^2 = 0.3185$$

Hence, $n = 4$, $p = 0.36$ and $P(X = 2) = 0.3185$.

Example 9.11.

Is it possible to have a binomial distribution with mean = 10 and standard deviation = 6?

Solution:

Here, $\text{Mean} = \mu = np = 10$

$S.D(X) = \sigma = \sqrt{npq} = 6$

$Var(X) = \sigma^2 = npq = (6)^2 = 36$

$10\,q = 36$ (since $np = 10$) or $q = \dfrac{36}{10} = 3.6$

This is impossible because q is a probability which can never exceed one. Hence the given values of mean and standard deviation are wrong.

Example 9.12.

Given $n = 5$, $P(X = 1) = 5/32$ and $P(X = 2) = 10/32$. Find $P(X = 0)$ and $P(X = 3)$. Also find coefficient of skewness.

Solution:

$$P[X = 1] = \binom{5}{1}pq^4 = 5pq^4 = \frac{5}{32} \quad \ldots\ldots \quad (1)$$

$$P[X = 2] = \binom{5}{2}p^2q^3 = 10p^2q^3 = \frac{10}{32} \quad \ldots\ldots \quad (2)$$

Dividing equation (2) by equation (1)

$$\frac{10\,p^2q^3}{5\,pq^4} = \frac{10/32}{5/32} \quad \text{or} \quad \frac{2p}{q} = 2 \quad \text{or} \quad 2p = 2q \quad \text{or} \quad p = q = \frac{1}{2}$$

$$P[X=0] = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$P[X=3] = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

Coefficient of skewness $= \dfrac{q-p}{\sqrt{npq}} = 0$, there is no skewness, thus the distribution is symmetrical.

9.8 PROPERTIES OF THE BINOMIAL DISTRIBUTION

Properties of the binomial distribution are listed in the following table.

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard Deviation	$\sigma = \sqrt{npq}$
Moment Coefficient of Skewness	$\gamma_1 = \sqrt{\beta_1} = \dfrac{q-p}{\sqrt{npq}}$
Moment Coefficient of Kurtosis	$\beta_2 = 3 + \dfrac{1-6pq}{npq}$

Note: Skewness is negative, zero or positive according as $p > 1/2$ or $p = 1/2$ or $p < 1/2$.

9.9 BINOMIAL FREQUENCY DISTRIBUTION

Let the n independent trials constitute one experiment and let this experiment be repeated N times. Then we expect x successes to occur $N\binom{n}{x}p^xq^{n-x}$ times. This will be called the expected frequency of x successes in N experiments and the possible number of successes together with the expected frequencies will be said to constitute the binomial (expected) frequency distribution. In practice, the observed frequencies will differ from the expected frequencies due to chance causes. For N sets, each of n trials the expected frequencies of 0, 1, 2, ..., n successes are given by the successive terms in the binomial expansion of $N(q+p)^n$, $q + p = 1$.

9.10 FITTING OF THE BINOMIAL DISTRIBUTION

Suppose there are some intelligent and some non-intelligent students in a big college but we do not know the percentage of intelligent students. A random sample of 6 students is selected and the number of intelligent students in the sample is counted. This number will take any value between 0 to 6. We repeat the sample a large number of times say 100 and each time note the number of intelligent students. The frequencies of 0, 1, 2, ..., 6 intelligent students as observed in this

experiment are, say, 14, 15, 25, 25, 10, 6, 5. The sample data can be written as below:

No. of intelligent students (X)	0	1	2	3	4	5	6
Frequency (f)	14	15	25	25	10	6	5

The sample mean $(\bar{X}) = \frac{\Sigma f X}{\Sigma f} = \frac{240}{100} = 2.4$. This observed data is also called experimental or actual data. From this data we want to estimate the proportion of intelligent students in the college. According to the theory of estimation, we find the sample mean = 2.4 and substitute it equal to np when n = 6. Thus 6p = 2.4 and p = 0.4. This value of p is called estimate of proportion of intelligent students in the college. When p = 0.4 then q = 1 – p = 1 – 0.4 = 0.6. Using n = 6, p = 0.4 and q = 0.6 we can find the probabilities of 0, 1, 2, ..., 6 intelligent students as given below. Each probability multiplied with 100 will give us expected frequencies of 0, 1, 2, ..., 6 intelligent students. These frequencies are called frequencies of the binomial distribution.

No. of intelligent students (X)	Probability $p(x) = \binom{6}{x}(0.4)^x (0.6)^{6-x}$	Expected Frequencies N.p(x)
0	$\binom{6}{0}(0.4)^0 (0.6)^6 = 0.046656$	4.67
1	$\binom{6}{1}(0.4)(0.6)^5 = 0.186624$	18.66
2	$\binom{6}{2}(0.4)^2 (0.6)^4 = 0.311040$	31.10
3	$\binom{6}{3}(0.4)^3 (0.6)^3 = 0.276480$	27.65
4	$\binom{6}{4}(0.4)^4 (0.6)^2 = 0.138240$	13.82
5	$\binom{6}{5}(0.4)^5 (0.6) = 0.036864$	3.69
6	$\binom{6}{6}(0.4)^6 (0.6)^0 = 0.004096$	0.41
Total	1	100

This procedure of finding expected frequencies of the binomial distribution is called fitting of the binomial distribution. It is basically the method of estimating

probability of success p(parameter of population) from the sample data and then use it to find the expected frequencies (most likely) of the given experiment.

Example 9.13.

The probability of male birth is equal to the probability of female birth. Out of 400 families with 4 children each, find the expected number of families with 0, 1, 2, 3 and 4 males.

Solution:

The binomial frequency distribution is

$$N.P[X = x] = N.b(x; n, p) = N\binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \ldots, n.$$

Here, $N = 400$, $n = 4$, $p = 1/2$, $q = 1 - p = 1/2$ and $x = 0, 1, 2, 3, 4$.

$$400\,P[X = 0] = 400\binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = 25 \text{ Families}$$

$$400\,P[X = 1] = 400\binom{4}{1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = 100 \text{ Families}$$

$$400\,P[X = 2] = 400\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = 150 \text{ Families}$$

$$400\,P[X = 3] = 400\binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = 100 \text{ Families}$$

$$400\,P[X = 4] = 400\binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = 25 \text{ Families}$$

Example 9.14.

Six dice are thrown 729 times. How many times do you expect at least three dice to show a five or a six?

Solution:

The binomial frequency distribution is

$$N.P[X = x] = N.b(x; n, p) = N\binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \ldots, n.$$

Here, $p = 1/3$, $q = 1 - p = 2/3$, $n = 6$ and $N = 729$.

Let the random variable X denote the number of dice to show a five or a six, then the possible values of X are 0, 1, 2, 3, 4, 5, 6. Therefore

$$729\,P[X = x] = 729\binom{6}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{6-x} \quad x = 0, 1, 2, \ldots, 6.$$

The expected number of times a least three dice will show a 5 or a 6 is

$$= 729\binom{6}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3 + 729\binom{6}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + 729\binom{6}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)$$

$$+ 729\binom{6}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0$$

$$= 729\left(\frac{160}{729}\right) + 729\left(\frac{60}{729}\right) + 729\left(\frac{12}{729}\right) + 729\left(\frac{1}{729}\right) = 233$$

Example 9.15.

Fit a binomial distribution to the following data, obtained by tossing a biased coin 5 times.

No. of heads	0	1	2	3	4	5	Total
Frequency	12	56	74	39	18	1	200

Solution: The value of p is not given. Let us first calculate p.

No. of heads (X)	0	1	2	3	4	5	Total
Frequency (f)	12	56	74	39	18	1	200
Product (f X)	0	56	148	117	72	5	398

Here, $\Sigma f = 200, \quad \Sigma fX = 398, \quad n = 5$

$$\bar{X} = np = \frac{\Sigma fX}{\Sigma f} = \frac{398}{200} = 1.99$$

or $5p = 1.99$ (since $n = 5$)

or $p = \dfrac{1.99}{5} = 0.398$ and $q = 1 - p = 1 - 0.398 = 0.602$

Hence the binomial distribution to be fitted to the data is

$$200P[X = x] = 200\binom{5}{x}(0.398)^x (0.602)^{5-x}$$

The expected or theoretical frequencies of 0, 1, 2, 3, 4, 5 successes are calculated as below:

No. of heads (X)	Probability $P[X = x]$	Expected Frequencies $N.P[X = x]$
0	$\binom{5}{0}(0.398)^0 (0.602)^5 = 0.079065$	15.81 or 16
1	$\binom{5}{1}(0.398)(0.602)^4 = 0.261360$	52.27 or 52
2	$\binom{5}{2}(0.398)^2 (0.602)^3 = 0.345586$	69.12 or 69
3	$\binom{5}{3}(0.398)^3 (0.602)^2 = 0.228477$	45.70 or 46
4	$\binom{5}{4}(0.398)^4 (0.602) = 0.075526$	15.10 or 15
5	$\binom{5}{5}(0.398)^5 (0.602)^0 = 0.009987$	2.00 or 2
Total	1	200

Example 9.16.

A certain event is believed to follow the binomial distribution. In 1024 samples of 5, the result was observed once 405 times and twice 270 times. Calculate the probabilities p and q.

Solution:

The binomial frequency distribution is

$$N.P[X = x] = N\binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \dots, n.$$

Here, $N = 1024,\ n = 5,\ x = 1$ and 2.

$$1024.P[X = 1] = 1024\binom{5}{1} pq^4 = 405$$

$$1024.P[X = 2] = 1024\binom{5}{2} p^2q^3 = 270$$

$$5120\,pq^4 = 405 \qquad \dots\dots (1)$$

$$10240\,p^2q^3 = 270 \qquad \dots\dots (2)$$

Dividing equation (2) by equation (1), we get

$$\frac{10240\,p^2q^3}{5120\,pq^4} = \frac{270}{405} \quad \text{or} \quad \frac{2p}{q} = \frac{2}{3}$$

$$\text{or } 6p = 2q \text{ or } 6p = 2(1-p) = 2 - 2p \text{ or } 6p + 2p = 2$$

$$\text{or } 8p = 2 \text{ or } p = \frac{2}{8} = 1/4 \text{ and } q = 1 - p = 3/4$$

Hence, $p = 1/4$ and $q = 3/4$

9.11 HYPERGEOMETRIC DISTRIBUTION

If the probability of a success is not constant, the hypergeometric distribution is particularly useful. Suppose we have **N** distinct objects divided into two classes, say a class of successes and a class of failures. Suppose there are **k** successes and **N – k** failures. We take at random a sample of size **n** and ask for the probability that exactly **x** of the objects in it are successes and **n – x** failures. The probability distribution of the hypergeometric random variable **x** is

$$P(X = x) = \frac{\begin{pmatrix}\text{No. of successes in the population} \\ \text{Taken x at a time}\end{pmatrix}\begin{pmatrix}\text{No. of failures in the population} \\ \text{Taken n} - \text{x at a time}\end{pmatrix}}{\begin{pmatrix}\text{Total population} \\ \text{Taken n at a time}\end{pmatrix}}$$

where, $x = 0, 1, 2, 3, \dots, n$ or k (whichever is less), $x \le k,\ n - x \le N - k$

or

$$P(X = x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad \begin{array}{l} x = 0, 1, 2, 3, \cdots, n \text{ or } k \text{ (whichever is less)} \\ x \le k,\ n - x \le N - k \end{array}$$

If $P(X = x)$ is the probability distribution, it should satisfy

(i) $P(X = x) \geq 0$ (ii) Sum of probabilities $= \sum_{x=0}^{n} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1$

Note: (1) Hypergeometric distribution has three parameters i.e; **N**, **n** and **k**.

(2) The number of successes in the sample **x** cannot exceed the number of successes in the population **k** or the sample size **n**. Thus the range of the hypergeometric random variable is limited to the sample size or to the number of successes in the population whichever is smaller.

(3) When **N** is large, hypergeometric distribution approaches the binomial distribution.

(4) Like the binomial distribution, the hypergeometric distribution may also be symmetrical or skewed. Whenever **p** = 1/2, the hypergeometric distribution will be symmetrical regardless of how large or small the value of **n**; however, when **p** ≠ 1/2, the distribution will be skewed.

9.12 HYPERGEOMETRIC EXPERIMENT

An experiment in which a random sample is selected without replacement from a known finite population and contains a relatively large proportion of the population, such that the probability of a success does not remain constant from trial to trial is called a hypergeometric experiment.

A hypergeometric experiment has the following properties (qualities):

(i) Each trial results in two outcomes which can be classified into success and failure.

(ii) The successive trials are dependent.

(iii) The probability of each outcome does not remain constant from trial to trial.

(iv) The experiment is repeated a fixed number of times.

When X be a random variable for the number of successes out of a sample of n items selected without replacement from a finite population of N items of the hypergeometric experiment is called a hypergeometric random variable. The probability distribution of the hypergeometric random variable is called the hypergeometric probability distribution.

9.13 PROPERTIES OF THE HYPERGEOMETRIC DISTRIBUTION

(1) The mean and variance of the hypergeometric distribution are as follows:

$$\mu = E(X) = \frac{nk}{N} \quad \text{and} \quad Var(X) = \sigma^2 = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$

If we set $\frac{k}{N} = p$, then the mean of the hypergeometric distribution coincides with the mean of the binomial distribution and the variance of the hypergeometric distribution is $\left(\frac{N-n}{N-1}\right)$ times the variance of the binomial distribution.

(2) When N is large, hypergeometric distribution approaches the binomial distribution.

Example 9.17.

An urn contains nine balls, five of them red and four blue. Three balls are drawn without replacement. Find the probability distribution of X = number of red balls drawn.

Solution:

The hypergeometric distribution is

$$P(X=x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, 3, \cdots, n \text{ or } k \text{ (whichever is less)}$$

Here N = 9, n = 3, k = 5, x = number of red balls then the possible values of x are 0, 1, 2, 3. Therefore

$$P(X=x) = h(x; 9, 3, 5) = \frac{\binom{5}{x}\binom{9-5}{3-x}}{\binom{9}{3}} \quad x = 0, 1, 2, 3.$$

$$P(X=0) = \frac{\binom{5}{0}\binom{9-5}{3-0}}{\binom{9}{3}} = \frac{\binom{5}{0}\binom{4}{3}}{\binom{9}{3}} = \frac{4}{84} \qquad P(X=1) = \frac{\binom{5}{1}\binom{9-5}{3-1}}{\binom{9}{3}} = \frac{\binom{5}{1}\binom{4}{2}}{\binom{9}{3}} = \frac{30}{84}$$

$$P(X=2) = \frac{\binom{5}{2}\binom{9-5}{3-2}}{\binom{9}{3}} = \frac{\binom{5}{2}\binom{4}{1}}{\binom{9}{3}} = \frac{40}{84} \qquad P(X=3) = \frac{\binom{5}{3}\binom{9-5}{3-3}}{\binom{9}{3}} = \frac{\binom{5}{3}\binom{4}{0}}{\binom{9}{3}} = \frac{10}{84}$$

Thus the hypergeometric probability distribution of red balls in tabular form is given as follows:

x	0	1	2	3	Total
P(X = x)	4/84	30/84	40/84	10/84	1

Example 9.18

There are seven people who work in an office. Of the seven, four would like to be transferred. If three people from this office are randomly selected for transfer, what is the probability that:

(i) All three will want to be transferred?

(ii) Two of the three will want to be transferred?

(iii) At least two will want to be transferred?

(iv) At most one will want to be transferred?

Solution:

The hypergeometric distribution is

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, 3, \ldots, n \text{ or } k \text{ (whichever is less)}$$

Here N = 7, n = 3, k = 4, x = number of people who wanted the transfer, then the possible values of x are 0, 1, 2 and 3. Therefore

$$P(X = x) = \frac{\binom{4}{x}\binom{7-4}{3-x}}{\binom{7}{3}} \quad x = 0, 1, 2, 3.$$

(i) $$P(X = 3) = \frac{\binom{4}{3}\binom{7-4}{3-3}}{\binom{7}{3}} = \frac{\binom{4}{3}\binom{3}{0}}{\binom{7}{3}} = \frac{4}{35} = 0.1143$$

(ii) $$P(X = 2) = \frac{\binom{4}{2}\binom{7-4}{3-2}}{\binom{7}{3}} = \frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} = \frac{18}{35} = 0.5143$$

(iii) $$P(X \geq 2) = P(X = 2) + P(X = 3) = \frac{18}{35} + \frac{4}{35} = \frac{22}{35} = 0.6286$$

(iv) $$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{\binom{4}{0}\binom{7-4}{3-0}}{\binom{7}{3}} + \frac{\binom{4}{1}\binom{7-4}{3-1}}{\binom{7}{3}}$$

$$= \frac{1}{35} + \frac{12}{35} = \frac{13}{35} = 0.3714$$

Example 9.19

A box contains ten items, seven of which are good and three are defective. Two items are selected (without replacement); compute the probability distribution for the number of defectives in the sample of two. Compute the mean and variance of this probability distribution. Is this mean equal to $\frac{nk}{N}$ and variance $\frac{nk(N-k)(N-n)}{N^2(N-1)}$.

Solution:

The hypergeometric distribution is

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, 3, \ldots, n \text{ or } k \text{ (whichever is less)}$$

Here N = 10, n = 2, k = 3, x = number of defective items, then the possible values of x are 0, 1 and 2.

Therefore

$$P(X=x)=\frac{\binom{3}{x}\binom{10-3}{2-x}}{\binom{10}{2}} \quad x=0,\ 1,\ 2.$$

$$P(X=0)=\frac{\binom{3}{0}\binom{10-3}{2-0}}{\binom{10}{2}}=\frac{\binom{3}{0}\binom{7}{2}}{\binom{10}{2}}=\frac{7}{15}$$

$$P(X=1)=\frac{\binom{3}{1}\binom{10-3}{2-1}}{\binom{10}{2}}=\frac{\binom{3}{1}\binom{7}{1}}{\binom{10}{2}}=\frac{7}{15}$$

$$P(X=2)=\frac{\binom{3}{2}\binom{10-3}{2-2}}{\binom{10}{2}}=\frac{\binom{3}{2}\binom{7}{0}}{\binom{10}{2}}=\frac{1}{15}$$

Thus the probability distribution in tabular form for the computation of mean and variance of X is given as follows:

x	$P(X = x) = p(x)$	$x\,p(x)$	$x^2\,p(x)$
0	7/15	0	0
1	7/15	7/15	7/15
2	1/15	2/15	4/15
	$\Sigma p(x) = 1$	$\Sigma x\,p(x) = 9/15$	$\Sigma x^2\,p(x) = 11/15$

$$\text{Mean} = E(X) = \Sigma x\,p(x) = \frac{9}{15} = 0.6$$

$$\text{Var}(X) = \sigma^2 = \Sigma x^2\,p(x) - [\Sigma x\,p(x)]^2 = \frac{11}{15} - \left(\frac{9}{15}\right)^2 = \frac{28}{75} = 0.3733$$

$$\frac{nk}{N} = \frac{(2)\,(3)}{10} = 0.6 \quad \text{and} \quad \frac{nk(N-k)(N-n)}{N^2(N-1)} = \frac{2(3)\,(7)\,(8)}{100(9)} = 0.3733$$

$$\text{Hence } E(X) = \frac{nk}{N} \quad \text{and} \quad \text{Var}(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$

SHORT DEFINITIONS

Bernoulli Trial

A trial that gives only two possible outcomes is called a Bernoulli trial.

Binomial Experiment

An experiment with n independent trials in which the outcomes can always be classified as either a success or a failure and the probability of success remains constant from trial to trial is called a binomial experiment.

Properties of a Binomial Experiment

(i) The experiment consists of **n** identical trials.

(ii) The trials are independent.

(iii) Each trial can result in one of only two possible outcomes, called success and failure.

(iv) The probability of success **p** is constant from trial to trial.

(v) The random variable X represents the number of successes in **n** trials.

Binomial Probability Distribution

A distribution that gives the probability of x successes for a fixed number of independent trials, where each trial must have two possible outcomes and the probability of a success is constant from trial to trial, is called binomial probability distribution.

or

A probability distribution showing the probability of x successes in **n** trials of a binomial experiment is called binomial probability distribution.

Binomial Probability Function

The function used to compute probabilities in a binomial experiment is called binomial probability function.

Binomial Formula

$$P(X = x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \cdots, n$$

where

n = Sample size $\qquad$ x = Number of successes

$n - x$ = Number of failures $\qquad$ p = Probability of a success

$q = 1 - p$ = Probability of a failure $\qquad$ $n! = n(n-1)(n-2)\cdots 3.2.1$

$0! = 1$ (by definition)

Hypergeometric Experiment

An experiment in which a random sample is selected without replacement from a known finite population and contains a relatively large proportion of the population, such that the probability of a success does not remain constant from trial to trial is called a hypergeometric experiment.

or

An experiment in which a random sample is selected without replacement from a finite population in such a way that each trial is a Bernoulli trial and probability of success does not remain constant on each trial, is called a hypergeometric experiment.

Properties of a Hypergeometric Experiment

(i) The experiment consists of n identical trials.

(ii) The successive trials are dependent.

(iii) Each trial can result in one of only two possible outcomes, called success and failure.

(iv) The probability of each outcome does not remain constant from trial to trial.

Hypergeometric Probability Distribution

Suppose a population consists of N items which are classified as k successes and N–k failures. If we select a sample of n items from the population without replacement in such a way that x successes are selected from k successes and n – x failures are selected from N – k failures. The probability distribution defined in this situation is a hypergeometric distribution which is given by

$$P(X = x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0, 1, 2, 3, \cdots, n \text{ or } k \text{ (whichever is less)}$$
$$x \le k,\ n - x \le N - k,\ n \le N.$$

MULTIPLE – CHOICE QUESTIONS

1. A Bernoulli trial has:
   (a) at least two outcomes (b) at most two outcomes
   (c) two outcomes (d) fewer than two outcomes
2. The two mutually exclusive outcomes in a Bernoulli trial are usually called:
   (a) success and failure (b) variable and constant
   (c) mean and variance (d) with and without replacement
3. Nature of the binomial random variable X is:
   (a) quantitative (b) qualitative
   (c) discrete (d) continuous

4. In a binomial probability distribution, the sum of probability of failure and probability of success is always:
   (a) zero (b) less than 0.5
   (c) greater than 0.5 (d) one
5. In a binomial experiment, the successive trials are:
   (a) dependent (b) independent
   (c) mutually exclusive (d) fixed
6. The parameters of the binomial distribution are:
   (a) n and p (b) p and q
   (c) np and nq (d) np and npq
7. The range of binomial distribution is:
   (a) 0 to n (b) 0 to ∞
   (c) -1 to $+1$ (d) 0 to 1
8. The mean and standard deviation of the binomial probability distribution are respectively:
   (a) np and npq (b) np and $\sqrt{npq}$
   (c) np and nq (d) n and p
9. In a binomial experiment with three trials, the variable can take:
   (a) 2 values (b) 3 values
   (c) 4 values (d) 5 values
10. The shape of the binomial probability distribution depends upon the values of its:
   (a) mean (b) variance
   (c) parameters (d) quartiles
11. In binomial distribution the number of trials are:
   (a) very large (b) very small
   (c) fixed (d) not fixed
12. In a binomial probability distribution, relation between mean and variance is:
   (a) mean < variance (b) mean = variance
   (c) mean > variance (d) difficult to tell
13. In binomial distribution when $n = 1$, then it becomes:
   (a) hypergeometric distribution (b) normal distribution
   (c) uniform distribution (d) Bernoulli distribution
14. The mean of a binomial distribution depends on:
   (a) number of trials (b) probability of success
   (c) probability of failure (d) number of trials and probability of success
15. The variance of a binomial distribution depends on:
   (a) number of trials (b) probability of success
   (c) probability of failure (d) all of the above

16. Which of the following is not property of a binomial experiment?
(a) Probability of success remains constant
(b) n is fixed
(c) Successive trials are dependent
(d) It has two parameters

17. The binomial probability distribution is symmetrical when:
(a) $p = q$ (b) $p < q$
(c) $p > q$ (d) $np > npq$

18. The binomial distribution is negatively skewed if:
(a) $p < 1/2$ (b) $p = 1/2$
(c) $p > 1/2$ (d) $p = 1$

19. In a binomial probability distribution, the skewness is positive for:
(a) $p < 1/2$ (b) $p = 1/4$
(c) $np = npq$ (d) $np = nq$

20. Which one of the following statements is false?
(a) Expected value of a constant is the constant.
(b) In a binomial distribution the standard deviation is always less than its variance.
(c) In a binomial probability distribution the mean is always greater than its variance.
(d) In binomial experiment the probability of success remains constant from trial to trial.

21. If a binomial probability distribution has parameters $(n, p) = (5, 0.6)$, the probability of $X = 3.5$ is:
(a) 0 (b) 1
(c) 0.6 (d) 0.4

22. In a binomial experiment $n = 4$, $P(X = 2) = 216/625$ and $P(X = 3) = 216/625$. $P(X = -2)$ is:
(a) 216/625 (b) 1
(c) 0 (d) difficult to tell

23. If $n = 6$ and $p = 0.9$, then the value of $P(X = 7)$ is:
(a) zero (b) less than zero
(c) more than zero (d) one

24. In a binomial probability distribution, coefficient of skewness $= (q - p)/\sqrt{npq} = 0$, it means that the distribution is:
(a) symmetrical (b) skewed to the left
(c) skewed to the right (d) highly skewed

25. For a binomial distribution with n = 10, p = 0.5, the probability of zero or more successes is:

(a) 1 (b) 0.5
(c) 0.25 (d) 0.75

26. In a binomial distribution, the mean, median and mode coincide when:

(a) $p < 1/2$ (b) $p > 1/2$
(c) $p \neq 1/2$ (d) $p = 1/2$

27. In which distribution, the probability success remains constant from trial to trial?

(a) Hypergeometric distribution (b) Binomial distribution
(c) Sampling distribution (d) Frequency distribution

28. In a binomial experiment when n = 5, the maximum number of successes will be:

(a) 0 (b) 2.5
(c) 4 (d) 5

29. In a binomial experiment when n = 10, the minimum number of successes will be:

(a) 0 (b) 5
(c) 10 (d) 11

30. If n = 10 and p = 0.6, then $P(X \geq 0)$ is:

(a) 0.5 (b) 0.6
(c) 1.0 (d) 1.2

31. A random variable X has a binomial distribution with n = 4, the standard deviation of X is:

(a) $4pq$ (b) $2\sqrt{pq}$
(c) $4p$ (d) $4(q+p)$

32. In a multiple choice test there are five possible answers to each of 20 questions. If a candidate guesses the correct answer each time, the mean number of correct answers is:

(a) 4 (b) 5
(c) 1/5 (d) 20

33. If three coins are tossed, the probability of two heads is:

(a) 1/8 (b) 3/8
(c) 2/3 (d) 0

34. A random variable X has binomial distribution with n = 8 and p = 1/2. The most probable value of X is:

(a) 2 (b) 3
(c) 4 (d) 5

35. The value of second moment about the mean in a binomial distribution is 36. The value of the standard deviation of a binomial distribution is:
(a) 36 (b) 6
(c) 1/36 (d) 1/6

36. For a binomial probability distribution, the expected frequency of x successes in N experiments is:
(a) $\binom{k}{x}\binom{N-k}{n-x}/\binom{N}{n}$ (b) $p^x(1-p)^{1-x}$
(c) $N\binom{n}{x}p^x q^{n-x}$ (d) $\binom{n}{x}p^x(1-p)^{n-x}$

37. In a binomial frequency distribution $100(1/5+4/5)^5$ The parameters **n** and **p** are respectively:
(a) (5, 1/5) (b) (1/5, 4/5)
(c) (100, 4/5) (d) (5, 4/5)

38. For a binomial frequency distribution $100(1/5+4/5)^5$, the mean is:
(a) 1/5 (b) 4/5
(c) 5 (d) 4

39. For a binomial distribution $(1/3+2/3)^{18}$, the standard deviation of the binomial distribution will be:
(a) 2 (b) 4
(c) 6 (d) 12

40. The hypergeometric distribution has:
(a) one parameter (b) two parameters
(c) three parameters (d) four parameters

41. The parameters of the hypergeometric distribution are:
(a) N, n, p (b) N, n, np
(c) N, n, k (d) n and p

42. Nature of the hypergeometric random variable is:
(a) continuous (b) discrete
(c) qualitative (d) quantitative

43. In hypergeometric distribution, the successive trials are
(a) independent (b) dependent
(c) very large (d) very small

44. In a hypergeometic distribution, the probability of success:
(a) remains constant from trial to trial
(b) does not remain constant from trial to trial
(c) equal to probability of failure
(d) less than probability of failure

45. If in a hypergeometric distribution $N=10$, $k=5$ and $n=4$; then the probability of failure is:
(a) 2 (b) 0.5
(c) 1 (d) 0.25

46. The range of hypergeometric distribution is:
(a) 0 to n
(b) 0 to k
(c) 0 to N
(d) 0 to n or k (whichever is less)

47. The number of trials in hypergeometric distribution is:
(a) not fixed
(b) fixed
(c) large
(d) small

48. The probability of a success changes from trial to trial in:
(a) binomial distribution
(b) hypergeometric distribution
(c) normal distribution
(d) frequency distribution

49. The mean of the hypergeometric distribution is:
(a) $\frac{nk}{N}$
(b) $\frac{Nk}{n}$
(c) $\frac{Nn}{k}$
(d) $\frac{n+k}{N}$

50. The standard deviation of the hypergeometric distribution is:
(a) $\frac{nk(k-N)(N-n)}{N^2(N-1)}$
(b) $\sqrt{\frac{nk(N-k)(N-n)}{N^2(N-1)}}$
(c) $\sqrt{\frac{nk(N-k)(n-N)}{N^2(N+1)}}$
(d) $\sqrt{\frac{nk(N-k)+(N-n)}{N^2(N-1)}}$

51. In hypergeometric probability distribution, the relation between mean and variance is:
(a) mean > variance
(b) mean < variance
(c) mean = variance
(d) mean = 2 variance

52. Which of the following is the property of hypergeometric experiment?
(a) p remains constant from trial to trial
(b) Successive trials are independent
(c) Sampling is performed without replacement
(d) n is not fixed

53. Hypergeometric distribution reduces to binomial distribution when:
(a) $N = n$
(b) $n \to \infty$
(c) $N \to \infty$
(d) $N < n$

54. In a hypergeometric distribution $N = 6$, $n = 4$ and $k = 3$, then the mean is equal to:
(a) 2
(b) 4
(c) 6
(d) 24

55. Given $N = 11$, $n = 5$, $k = 7$; $P(X \geq 1)$ equals:
(a) 1
(b) 1/66
(c) 65/66
(d) none of the above

56. Given $N = 12$, $n = 5$, $k = 4$; $P(X \leq 4)$ equals:
(a) less than one
(b) exactly one
(c) more than one
(d) between 0.5 and 1

Answers

1.	(c)	2.	(a)	3.	(c)	4.	(d)	5.	(b)	6.	(a)	7.	(a)	8.	(b)
9.	(c)	10.	(c)	11.	(c)	12.	(c)	13.	(d)	14.	(d)	15.	(d)	16.	(c)
17.	(a)	18.	(c)	19.	(a)	20.	(b)	21.	(a)	22.	(c)	23.	(a)	24.	(a)
25.	(a)	26.	(d)	27.	(b)	28.	(d)	29.	(a)	30.	(c)	31.	(b)	32	(a)
33.	(b)	34.	(c)	35.	(b)	36.	(c)	37.	(d)	38.	(d)	39.	(a)	40.	(c)
41.	(c)	42.	(b)	43.	(b)	44.	(b)	45.	(b)	46.	(d)	47.	(b)	48.	(b)
49.	(a)	50.	(b)	51.	(a)	52.	(c)	53.	(c)	54.	(a)	55.	(a)	56.	(b)

SHORT QUESTIONS

Q.1 Define a Bernoulli trial.

Q.2 Define binomial experiment.

Q.3 What properties must an experiment possess to be classed as a binomial experiment? or State the properties of a binomial experiment.

Q.4 What is a binomial experiment and what are its properties?

Q.5 Define binomial probability distribution.

Q.6 Define binomial probability function.

Q.7 Describe the conditions of a binomial experiment.

Q.8 State the formula used to calculate binomial probabilities?

Q.9 Which formulae should be used to find the mean, variance, standard deviation and coefficient of variation of a binomial random variable?

Q.10 Write down the properties of the binomial probability distribution.

Q.11 What is a binomial distribution and what are its properties?

Q.12 What are the parameters of a binomial distribution?

Q.13 Describe the binomial distribution.

Q.14 Describe the binomial frequency distribution.

Q.15 If X is a binomial random variable with $n = 4$ and $P(X = 2) = 3P(X = 3)$. Find p

Ans. 1/3

Q.16 A random variable X is believed to follow a binomial distribution with $b(x, 5, p)$. If $P(X = 0) = 243/1024$, find $P(X = 3)$.

Ans. 0.0879

Q.17 In a binomial experiment $n = 5$, $P(X = 1) = 5/32$ and $P(X = 2) = 10/32$. Find $P(X = 5)$.

Ans. 1/32

Q.18 A random variable X has a binomial distribution with $n = 5$ and $p = 0.2$. Find $P(X = 2)$.

Ans. 0.2048

Q.19 For a binomial distribution with $n = 10$ and $p = 0.5$. Find the probability of 5 successes.

Ans. 0.2461

Q.20 If X is a binomial random variable with mean = 2.4 and variance = 0.96. Find $P(X = 0)$.

Ans. 0.0256

Q.21 In a binomial distribution consisting of 5 independent trials, probability of 2 and 3 successes are 0.2048 and 0.0512 respectively. Find the value of variance.

Ans. 0.8

Q.22 If X is a random variable with a binomial distribution and its mean equals 3.2 and its variance equals 1.92. Find the probability that all successes are observed.

Ans. 0.0007

Q.23 In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the expected value.

Ans. 1

Q.24 In a binomial distribution mean = 2.4 and standard deviation = 1.2. Find the value of n.

Ans. 6

Q.25 A random variable X has binomial distribution with $E(X) = 2.4$ and $p = 0.3$. Find the standard deviation of X.

Ans. 1.30

Q.26 If a binomial probability distribution has parameters $(n, p) = (20, 0.6)$. Find the expected number of successes and standard deviation.

Ans. (12; 2.19)

Q.27 If X is a binomial random variable with $n = 10$ and $p = 0.4$. Find S.D.$(3 - 2X)$.

Ans. 3.1

Q.28 If X is a binomial random variable with mean = 1.44 and standard deviation = 0.96. Find $P(X = 2)$.

Ans. 0.3185

Q.29 Given $n = 5$, $P(X = 0) = 1/32$ and $P(X = 3) = 10 / 32$. Find coefficient of skewness.

Ans. zero

Q.30 Discuss the statement that in a binomial distribution mean = 5 and standard deviation = 5.

Ans. Impossible because $q = 5$.

Q.31 In a multiple choice test there are five possible answers to each of 20 questions. If a candidate guesses the correct answer each time, find standard deviation of correct answers.

Ans. 1.79

Q.32 It is estimated that 1 in 20 students are left-handed. What sample size should be taken to ensure that the expected number of left-handed students in the sample is 3? Also find the value of parameter σ.

Ans. $n = 60$ and $\sigma = 1.69$

Q.33 The mean and second moment about the mean in a binomial distribution are 10 and 6 respectively. Find the value of **n**.

Ans. 25

Q.34 If the probability that a day is fine is 0.6. Find the expected number and standard deviation of the fine days in a week.

Ans. (4.2, 1.30)

Q.35 If X is a binomial random variable with $n = 10$ and $p = 0.4$. Find mean and variance of $Y = \frac{3 - 2X}{5}$.

Ans. -1 and 0.384

Q.36 Discuss the statement that in a binomial distribution mean = 5 and variance = 5.

Ans. Since q is a probability of failure which is equal to one. In this case $p = 0$ and n is undefined, therefore the given statement is incorrect.

Q.37 5 % of items in a large batch are defective. If 50 items are selected at random, find the probability that at least one will be defective.

Ans. 0.9231

Q.38 If X is binomially distributed with $n = 10$ and $p = 0.4$.Find the variance of $Y = (X - 10)/6$.

Ans. 0.0667

Q.39 A supermarket stocks eggs are in boxes of six, and 10 % of the eggs are cracked. Assuming that the cracked eggs are distributed at random, compute the probability that a customer will find that the first box he chooses contains no cracked egg.

Ans. 0.5314

Q.40 In a binomial distribution with $n = 5$, $P(X = 0) = P(X = 1)$. Find the variance.

Ans. 25/36

Q.41 It is known that 40 % of the defective parts produced in a certain manufacturing process can be made satisfactory by rework. Find the probability that in a batch of six such defective parts none of them can be reworked.

Ans. 0.0467

Q.42 If X is a binomial random variable with $n = 25$ and $p = 0.2$. Find $P(X < \mu - 2\sigma)$.

Ans. 0.0038

Q.43 Define hypergeometric experiment.

Q.44 Write down the properties of hyper ometric experiment.

Q.45 Define hypergeometric distribution.

Q.46 State the formula of hypergeometric distribution.

Q.47 Write down the equation of the hypergeometric distribution. Also write its mean and variance.

Q.48 Write down parameters, mean and variance of the hypergeometric distribution.

Q.49 Write down the properties of the hypergeometric distribution.

Q.50 Differentiate between binomial and hypergeometric distributions.

Q.51 Given $N = 10$, $n = 4$ and $k = 3$. Find $P(X = 1)$.

Ans. 0.5

Q.52 Given $N = 10$, $n = 2$ and $k = 3$. Find $P(X = 0)$.

Ans. 0.47

Q.53 Given $N = 10$, $n = 4$ and $k = 5$. Find $E(X)$.

Ans. 2

Q.54 Given $N = 40$, $n = 5$ and $k = 8$. Find $Var(X)$.

Ans. 0.72

Q.55 Given $N = 10$, $n = 4$ and $k = 7$. Find $S.D.(X)$.

Ans. 0.75

Q.56 If X is a hypergeometric random variable with $N = 40$, $n = 5$ and $k = 4$. Find the values of mean and variance.

Ans. 0.5 and 0.4

Q.57 A random variable X has hypergeometric distribution with $N = 11$, $n = 5$ and $k = 7$. Find $P(X \leq 1)$.

Ans. 7/462.

Q.58 If 13 cards are chosen at random (without replacement) from an ordinary deck of 52 cards. Find the probability that 6 are picture cards.

Ans. 0.0271

EXERCISES

Q.1 An event has probability p = 3/5. Find complete binomial distribution for n = 5.

Ans. 32/3125, 240/3125, 720/3125, 1080/3125, 810/3125, 243/3125

Q.2 A and B play a game in which A's chances of winning are 2/3. A series of 5 games is played. Find the probability that:

(i) A will win 3 games (ii) A will win at least 3 games

(iii) A will win at most 3 games.

Ans. (i) 80/243 (ii) 192/243 (iii) 131/243

Q.3 A fair coin is tossed four times. What is the probability of getting:

(i) exactly one head (ii) exactly two heads (iii) at least two heads

(iv) at most three heads (v) between one and three heads inclusive.

Ans. (i) 1/4 (ii) 3/8 (iii) 11/16 (iv) 15/16 (v) 7/8

Q.4 What is the probability of obtaining (i) 2 sixes, (ii) 3 sixes, (iii) at least 4 sixes and (iv) at most 3 sixes when a perfect cubical die is thrown 5 times?

Ans. (i) 1250/7776 (ii) 250/7776 (iii) 26/7776 (iv) 7750/7776

Q.5 A machine produces parts which are graded A or B according to finish and the proportion of A to B is well established at 2:1. Calculate the probabilities of obtaining 0, 1, 2, 3, 4, 5 parts of finish A in any sample of 5.

Ans. 1/243, 10/243, 40/243, 80/243, 80/243, 32/243

Q.6 If 20 % of the bolts produced by a machine are defective, determine the probability that of 5 bolts chosen at random.

(i) 2 bolts are defective (ii) at least 3 bolts are defective.

Ans. (i) 640/3125 (ii) 181/3125

Q.7 Team A has probability $\frac{2}{3}$ of winning when ever it plays. If A plays 4 games find the probability that A wins: (i) exactly 2 games (ii) at least one game

(iii) more than half of the games.

Ans. (i) 8/27 (ii) 80/81 (iii) 16/27

Q.8 A rat has five choices of alternate routes in order to reach the food box. If it is true that for each choice the odds are two to one in favour of the correct pathway, what is the probability that the rat will make:

(i) all of its choices correctly

(ii) none of its choices correctly

(iii) at least one of its choices correctly

(iv) at least two of its choices correctly.

Ans. (i) 32/243 (ii) 1/243 (iii) 242/243 (iv) 232/243

Q.9 The probability that a patient recovers from a disease is 0.8. Suppose 5 people are known to have contracted the disease. Assuming independence find the probability that:

(i) exactly 4 recover (ii) at least one recover

(iii) at least one but not more than three recover (iv) at most three recover.

Ans. (i) 0.4096 (ii) 0.99968 (iii) 0.2624 (iv) 0.26272

Q.10 The probability that a college student can pass a subject is $\frac{3}{5}$. Find the chance that out of 5 students: (i) at least 3 will fail (ii) at most 3 will pass.

Ans. (i) 992/3125, (ii) 2072/3125.

Q.11 The experience of births at a maternity hospital over a long period of time indicates that the babies born are boys and girls in equal number. If in a particular week 8 pregnant women are admitted to the hospital, determine the probability that: (i) there will be an equal number of boys and girls

(ii) there will be at least 7 girls.

Ans. (i) 0.2734 (ii) 0.0352

Q.12 The record of a heart specialist shows that 75 % of the patients he operates survive. If on a particular day he operates 6 patients, find the probability that:

(i) fewer than 4 patients will survive (ii) exactly 4 patients will survive

(iii) at least 5 patients will survive.

Ans. (i) 0.1694 (ii) 0.2966 (iii) 0.5339

Q.13 In a certain district, the need for money to buy drugs is given as the reason for 75% of all thefts. What is the probability that exactly 2 of the next 4 theft cases reported in this district resulted from the need for money to buy drugs?

Ans. 0.2109

Q.14 A six faced die is so biased that it is twice as likely to show an even number as an odd number. If it is thrown five times, find the probability that it shows:

(i) no even number (ii) all even numbers.

Ans. (i) 1/243 (ii) 32/243

Q.15 What is the probability of getting a total of 9:

(i) twice (ii) at most twice (iii) at least twice in 4 throws of a pair of dice?

Ans: (i) 0.0585 (ii) 0.9950 (iii) 0.0636

Q.16 If a cadet succeeds 4 times out of 5 in shooting in bull's eye, what is the probability that he will miss not more than 3 times in 10 attempts.

Ans. 0.8791

Q.17 Find mean, variance and standard deviation of binomial $(q + p)^4$.

Ans. Mean = 4p, Variance = 4pq, S.D. = $\sqrt{4pq}$

Q.18 In a binomial distribution mean = 2.4 and S.D. = 1.2. Find the value of p and n.

Ans. p = 0.4, n = 6

Q.19 A random variable X has a binomial distribution with $E(X) = 2.4$ and $p = 0.3$. Find n and standard deviation of X.

Ans. $n = 8$, $S.D(X) = 1.2961$

Q.20 If X is binomially distribution with $n = 10$ and $p = 0.4$ then find mean and variance of $Y = \frac{X - 10}{6}$.

Ans. $E(Y) = -1, Var(Y) = 0.0667$

Q.21 Is it possible to have a binomial distribution with (simply write "yes" or "no"):

(i)	Mean = 5 and variance = 2.5	(ii)	$P(X = -2) = 0.1467$
(iii)	$P(X \leq 1) = 0.35$	(iv)	Three parameters viz. n, p and q
(v)	$P(X = 1) = 0.18$	(vi)	$P(X \geq 1) = 1.27$
(vii)	$P(X = 3.8) = 0.29$	(viii)	$P(1 \leq X \leq 4) = 0.79$

Ans. (i) Yes (ii) No (iii) Yes (iv) No (v) Yes

(vi) No (vii) No (viii) Yes

Q.22 Is it possible to have a binomial distribution with mean 15 and standard deviation 3?

Ans. Yes

Q.23 Find the mean and variance of binomial distribution $(q + p)^3$.

Ans. Mean = 3p, Variance = 3pq

Q.24 A binomial distribution has mean = 1.44 and variance = 0.9216. Find probability that X is greater than 2.

Ans. 0.136235

Q.25 If a coin is tossed 5 times and the probability that it shows 3 heads is twice the probability that it shows one tail, then find q and p.

Ans. $p = q = 1/2$.

Q.26 Find the mean and variance of the binomial probability distribution $(q + p)^n$.

Ans. Mean = np, Variance = npq

Q.27 A random variable X has a binomial distribution with $E(X) = 2$ and $Var(X) = 24/13$. Find the values of n and p and $P(X = 2)$.

Ans. $n = 26, p = \frac{1}{13}$, $P(X = 2) = 0.28165$,

Q.28 If X is a binomial random variable with $n = 10$ and $p = 0.4$, then find its mean, variance, standard deviation, $E(2X + 3)$, $Var(2 - 3X)$ and $S.D(3 - 2X)$.

Ans. $E(X) = 4$, $Var(X) = 2.4$, $S.D(X) = 1.55$, $E(2X + 3) = 11$,

$Var(2 - 3X) = 21.6$, $S.D(3 - 2X) = 3.1$.

Q.29 Given that X is a binomial random variable with parameters $n = 1$ and p; prove that $E(X) = p$ and $S.D(X) = \sqrt{pq}$.

Q.30 A random variable X has binomial distribution with $n = 5$ and $p = 0.4$. Find (i) the most probable value of X (ii) the expected number of successes.

Ans: (i) 2 (ii) 2

Q.31 For a binomial distribution with $n = 4$ and $p = 1/4$. Find the complete binomial distribution. Also find the mean and variance of the probability distribution you have obtained. Is this mean equal to np and variance npq?

Ans: $E(X) = 1$ $Var(X) = 0.75$

Q.32 Out of 2000 families with 4 children each, how many would you expect to have:
(i) at least one boy (ii) two boys (iii) one or two girls (iv) no girl
Assuming equal probabilities for boys and girls.

Ans. (i) 1875 (ii) 750 (iii) 1250 (iv) 125

Q.33 Five coins are tossed 96 times:

(i) Construct the theoretical frequency table 0, 1, 2, 3, 4, 5 heads.
(ii) Find the expected number of times of getting at least 3 heads.

Ans. (i) 3, 15, 30, 30, 15, 3 (ii) 48

Q.34 Five dice are thrown 243 times. How many times do you expect at least two dice to show a three or four.

Ans. 131

Q.35 Determine the following:

(i) If $n = 4$, $N = 10$, and $k = 5$, then find $P(X = 3)$.
(ii) If $n = 4$, $N = 6$, and $k = 3$, then find $P(X = 1)$.
(iii) If $n = 5$, $N = 12$, and $k = 3$, then find $P(X > 1)$.
(iv) If $n = 3$, $k = 3$, and $N = 10$, then find $P(1 \le X < 3)$.

Ans. (i) 0.2381 (ii) 0.2 (iii) 0.3636 (iv) 0.7

Q.36 Five balls are drawn from a box containing 7 red and 4 blue balls. If X denotes the number of red balls drawn from the box, then obtain the probability distribution of X. Compute the mean and variance of this distribution.

Ans.

x	1	2	3	4	5
p(x)	1/66	12/66	30/66	20/66	3/66

Mean = 3.1818,
Variance = 0.6942

Q.37 A box contains ten items, seven of which are good and three are defective. A sample of four items is to be selected. Compute the probability distribution for the numbers of defectives. Also compute the mean and standard deviation of this probability distribution.

Ans:

x	0	1	2	3	Total
p(x)	35/210	105/210	63/210	7/210	1

$\mu = 1.2$
$\sigma = 0.75$

Q.38 A state lottery is conducted in which six winning numbers are selected from a total of 54 numbers. What is the probability that if six numbers are randomly selected,

(i) four numbers will be winning numbers?

(ii) three numbers will be winning numbers?

(iii) none of the numbers will be winning numbers?

(iv) at least one number will be winning numbers?

Ans. (i) 0.0007 (ii) 0.0134 (iii) 0.4751 (iv) 0.5249

Q.39 What is the probability of drawing five cards from a deck and getting:

(i) three spades (ii) at most two spades (iii) at least three spades?

Ans. (i) 0.0815 (ii) 0.9072 (iii) 0.0928

Q.40 From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fire, what is the probability that

(i) all 4 will fire? (ii) at least 3 will fire?

(iii) at most 2 will not fire? (iv) at least 2 will not fire?

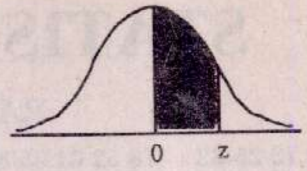
Ans. (i) 0.1667 (ii) 0.6667 (iii) 0.9667 (iv) 0.3333

STATISTICAL TABLES

RANDOM NUMBERS

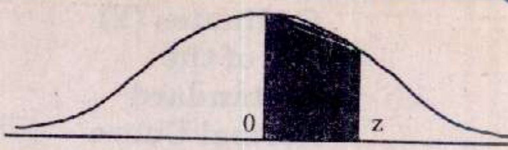
10 09 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17	39 29 27 49 45
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02	00 82 29 16 65
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64	35 08 03 36 06
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97	04 43 62 76 59
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77	12 17 17 68 33
66 06 57 47 17	34 97 27 68 50	36 69 73 61 70	65 81 33 98 85	11 19 92 91 70
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39	23 40 30 97 32
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47	18 62 38 85 79
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09	83 49 12 56 24
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44	35 27 38 84 35
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33	50 50 07 39 98
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01	52 77 56 78 51
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 53 06 10	68 71 17 78 17
88 68 54 02 00	86 50 74 84 01	36 76 66 79 51	90 36 47 64 93	29 60 91 10 62
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68	23 47 83 41 13
65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86	40 21 81 65 44
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53	14 38 55 37 63
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37	96 28 60 26 55
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90	94 40 05 64 18
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22	54 38 21 45 98
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23	37 08 92 00 48
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40	42 05 08 23 41
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81	22 22 20 64 13
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39	28 70 72 58 15
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82	07 20 73 17 90
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93	42 58 26 05 27
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18	33 21 15 94 66
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92	92 92 74 59 73
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59	25 70 14 66 70
23 52 37 83 17	73 20 8 98 37	68 93 59 14 16	26 25 22 96 63	05 52 28 25 62
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35	65 33 71 24 72
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91	23 28 72 95 29
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24	90 10 33 93 33
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92	78 56 52 01 06
46 05 88 52 36	01 39 00 22 86	77 28 14 40 77	93 91 08 36 47	70 61 74 29 41
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57	85 39 41 18 38
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23	97 11 89 63 38
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06	84 99 28 52 07
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33	20 82 66 95 41
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56	05 01 45 11 76

AREAS UNDER THE STANDARD NORMAL CURVE from 0 to z (4-places of decimals)



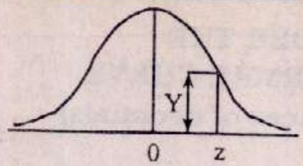
z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4[illegible]72	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4[illegible]79	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

AREA UNDER THE STANDARD NORMAL CURVE
From 0 to z (5-places of decimals)



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03585
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	15542	.15910	.16276	.16640	.17003	.17364	.17726	.18082	.18430	.18793
0.5	.19146	.19497	.19847	.20184	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23566	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29388	.29673	.29955	.30234	.30511	.30785	.33646	.33891
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34850	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.28298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40220	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42786	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44170	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.56495	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47616	.47679
2.0	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49243	.49361
2.5	.49379	.49393	.49413	.49430	.49446	.49461	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49861
3.0	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49897	.49897	.49800
3.1	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929
3.2	.49931	.49934	.49936	.49938	.49940	.49942	.49944	.49946	.49948	.49950
3.3	.49952	.49953	.49955	.49957	.49958	.49960	.49961	.49962	.49964	.49965
3.4	.49966	.49968	.49969	.49970	.49971	.49972	.49973	.49974	.49975	.49976
3.5	:49977									
3.6	.49984									
3.7	.49989									
3.8	.49993									
3.9	.49995									
4.0	.49997									

Ordinates (Y) of the Standard Normal Curve at z



z	0	1	2	3	4	5	6	7	8	9
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	0.3970	0.3965	0.3961	0.3956	0.3951	0.3945	0.3939	0.3932	0.3925	0.3918
0.2	0.3910	0.3902	0.3894	0.3885	0.3876	0.3867	0.3857	0.3847	0.3836	0.3825
0.3	0.3814	0.3802	0.3790	0.3778	0.3765	0.3752	0.3739	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	0.3637	0.3621	0.3605	0.3589	0.3572	0.3555	0.3538
0.5	0.3521	0.3503	0.3485	0.3467	0.3448	0.3429	0.3410	0.3391	0.3372	0.3352
0.6	0.3332	0.3312	0.3292	0.3271	0.3251	0.3230	0.3209	0.3187	0.3166	0.3144
0.7	0.3123	0.3101	0.3079	0.3056	0.3034	0.3011	0.2989	0.2966	0.2943	0.2920
0.8	0.2897	0.2874	0.2850	0.2827	0.2803	0.2780	0.2756	0.2732	0.2709	0.2685
0.9	0.2661	0.2637	0.2613	0.2589	0.2565	0.2541	0.2516	0.2492	0.2468	0.2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	0.2179	0.2155	0.2131	0.2107	0.2083	0.2059	0.2036	0.2012	0.1989	0.1965
1.2	0.1942	0.1919	0.1895	0.1872	0.1849	0.1826	0.1804	0.1781	0.1758	0.1736
1.3	0.1714	0.1691	0.1669	0.1647	0.1626	0.1604	0.1582	0.1561	0.1539	0.1518
1.4	0.1497	0.1476	0.1456	0.1435	0.1415	0.1394	0.1374	0.1354	0.1334	0.1315
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127
1.6	0.1109	0.1092	0.1074	0.1057	0.1040	0.1023	0.1006	0.0989	0.0973	0.0957
1.7	0.0940	0.0925	0.0909	0.0893	0.0878	0.0863	0.0848	0.0833	0.0818	0.0804
1.8	0.0790	0.0775	0.0761	0.0748	0.0734	0.0721	0.0707	0.0694	0.0681	0.0669
1.9	0.0656	0.0644	0.0632	0.0620	0.0608	0.0596	0.0584	0.0573	0.0562	0.0551
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449
2.1	0.0440	0.0431	0.0422	0.0413	0.0404	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	0.0347	0.0339	0.0332	0.0325	0.0317	0.0310	0.0303	0.0297	0.0290
2.3	0.0283	0.0277	0.0270	0.0264	0.0258	0.0252	0.0246	0.0241	0.0235	0.0229
2.4	0.0224	0.0219	0.0213	0.0208	0.0203	0.0198	0.0194	0.0189	0.0184	0.0180
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139
2.6	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.7	0.0104	0.0101	0.0099	0.0096	0.0093	0.0091	0.0088	0.0086	0.0084	0.0081
2.8	0.0079	0.0077	0.0075	0.0073	0.0071	0.0069	0.0067	0.0065	0.0063	0.0061
2.9	0.0060	0.0058	0.0056	0.0055	0.0053	0.0051	0.0050	0.0048	0.0047	0.0046
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034
3.1	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	0.0025	0.0025
3.2	0.0024	0.0023	0.0022	0.0022	0.0021	0.0020	0.0020	0.0019	0.0018	0.0018
3.3	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013	0.0013
3.4	0.0012	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010	0.0010	0.0009	0.0009
3.5	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.0006
3.6	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0004
3.7	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003	0.0003
3.8	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
3.9	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001

Critical Values for Student's t Distribution

df v	Level of significance for one-tailed test					
	0.10	0.05	0.025	0.01	0.005	0.0005
	Level of significance for two-tailed test					
	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Critical Values for the Chi-square (χ^2) Distribution

df	α						
v	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	6.251	7.815	9.348	9.837	11.345	12.838	16.268
4	7.779	9.488	11.143	11.668	13.277	14.860	18.465
5	9.236	11.070	12.832	13.388	15.086	16.750	20.517
6	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	12.017	14.067	16.013	16.622	18.475	20.278	24.322
8	13.362	15.507	17.535	18.168	20.090	21.955	26.125
9	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	19.812	22.362	24.736	25.472	27.688	29.819	34.528
14	21.064	23.685	26.119	26.873	29.141	31.319	36.123
15	22.307	24.996	27.488	28.259	30.578	32.801	37.697
16	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	24.769	27.587	30.191	30.995	33.409	35.718	40.790
18	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	27.204	30.144	32.852	33.687	36.191	38.582	43.820
20	28.412	31.410	34.170	35.020	37.566	39.997	45.315
21	29.615	32.671	35.479	36.343	38.932	41.401	46.797
22	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	34.382	37.652	40.646	41.566	44.314	46.928	52.620
26	35.563	38.885	41.923	42.856	45.642	48.290	54.052
27	36.741	40.113	43.194	44.140	46.963	49.645	55.476
28	37.916	41.337	44.461	45.419	48.278	50.993	56.893
29	39.087	42.557	45.722	46.693	49.588	52.336	58.302
30	40.256	43.773	46.979	47.962	50.892	53.672	59.703